



RECEIVED

SEP 18 2002

TECH CENTER 1600/2900

Inversion start site

|

860

ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
+-----+-----+-----+-----+-----+ 900
TAATATTCCTTTCTTTATTGCGTTACCTGTTACCCAC (41)
Y K G K R K * R N G Q V V

901

AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCAAAACCAAAGTGAGGTAGA
-----+-----+-----+-----+-----+-----+ 960
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTGTGGGGTTTGGTTCACTCCATCT (101)
K L * T Q V C T I I R N T P K P K * G R

961

AATAGCATGAGAAGCCGTGTTGATGTTAATTAAATT
-----+-----+-----+----- 996
TTATCGTACTCTCGGCACAAACTACAATTAA (137)
N S M R S R V * C * L I

The inversion sequence of the apo-dystrophin-4 cDNA (SEQ ID NO: 1)

Figure 1



RECEIVED

SEP 18 2002

TECH CENTER 1600/2900

Inversion start site

|

850

TAAAGAAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
+-----+-----+-----+-----+-----+ 900
ATTTCTTCTTAATTTCTTTCTTTATTGCGTTACCTGTTACCCAC (51)
* R K N Y K G K R K * R N G Q V V

901

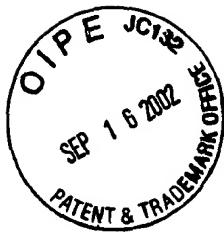
AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAACCAAAGTGAGGTAGA
+-----+-----+-----+-----+-----+-----+ 960
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTGTGGGTTTGGTTCACTCCATCT (111)
K L * T Q V C T I I R N T P K P K * G R

961

AATAGCATGAGAAGCCGTGTTGATGTTAATTAAATT
+-----+-----+-----+-----+ 996
TTATCGTACTCTCGGCACAAACTACAATTAAATTAA (147)
N S M R S R V * C * L I

The inversion sequence of the apo-dystrophin-4 cDNA plus a 10 base-pair region 5' to the start of the inversion sequence (SEQ ID NO: 1A, included in SEQ ID NO: 2).

Figure 1A



Start at 710

|

AACAATGGCAG

+-----+ 720
TTGTTACCGTC (11)

Q W Q

721 GTTTTACACGTCTATGCAATTGTACAAAAAGTTATAAGAAAACATGTAAAATCTG 780
-----+-----+-----+-----+-----+-----+
CAAAATGTGCAGATACGTTAACATGTTTTCAATATTCTTTGATGTACATTTAGAAC (71)
V L H V Y A I V Q K S Y K K T T C K I L

781 ATAGCTAAATAACTGCCATTCTTTATATGGAACGCATTTGGGTTGTTAAAAATTAA 840
-----+-----+-----+-----+-----+-----+
TATCGATTTATTGAACGGTAAAGAAATATACCTTGCCTAAACCCAAACAATTTTAAAT (131)
I A K * L A I S L Y G T H F G L F K N L
inversion start site

841 TAACAGTTATAAAAGAAAGAATTATAAAGAAAAAGAAAACGCAATGGACAAGTGGTG 900
-----+-----+-----+-----+-----+
ATTGTCATATTCTTTCTTAAATATTCTTTCTTTATTGCGTTACCTGTTACCCAC (191)
* Q L * R K N Y K G K R K * R N G Q V V

901 AAGCTGTGAACTCAGGTGTGCACAATTACAGGAACACCCAAAACCAAAGTGAGGTAGA 960
-----+-----+-----+-----+-----+
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTGTGGGTTTGGTTCACTCCATCT (251)
K L * T Q V C T I I R N T P K P K * G R

961 AATAGCATGAGAAGCCGTGTTGATGTTAATTAAATT 996
-----+-----+-----+-----
TTATCGTACTCTCGGCACAAACTACAATTAAATTAA (287)
N S M R S R V * C * L I

The inversion sequence of the apo-dystrophin-4 cDNA plus the upstream 150 bp from the start of the inversion at 860 to the Hpa I enzyme site (SEQ ID NO: 1B, included in SEQ ID NO: 2)

Figure 1B



1 GTGGTTGATTGATAGTAAAAAAATGTCGTTAACAGTAGAGAGTAAGTAATCAAT 60
 CACCAAACAACTATCATTTTTACAAGCAATTATGTTCATCTCTCATTCATTAGTTA
 V V * L I V K K M F V N T S R E * V I N

 61 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCTGTTCTGAT 120
 GTTAGTGAGTATCGGTTCCACCTTTCTACATAGGGTAGTACCTTATAAGGACAAGACTA
 Q S L I A K V E K M Y P I M E Y S C S D

 121 AGAAATCTTGTGCTTATCTATGGAATTCTTTGATATATATTACATTGGAACCTGAAT 180
 TCTTTAGAACACGAATAGATACCTTAAGAAAACATATATAATGTAACCCTGGACTTA
 R N L V L I Y G I L L I Y I Y I G N L N

 181 GTAGCTTGACATTTCCATGTAACACCAAGTAGCCTGATCCAACATTAAGCTGATACTA 240
 CATCGAACTGTAAAAGGTACATTGTGGTCATCGGACTAGGTTGTAATTGACTATGAT
 V A * H F S M * T P V A * S N I K L I L

 241 ACAAAACACGTGTAATGGCTTCATTAATAAGGCTTGCTTCTGGAAACTGGTAAA 300
 TGTTTGGCACATTACCGAAGTAATTATTCCGAAACGAAGAAGGACCTTGACCACTTT
 T N N V * W L H * * G F A S S W K L V K

 301 AATCAAACCTTGTGTAACCCCTCGATGCAGCTCTGTGTTCTTCACCCAGAAATG 360
 TTAGTTGGAACACACATGTGGGAGCTACGTCGAAGACACAAACAGAAGTGGTCTTAC
 N Q T L L C T P S M Q L L C C L H P E M

The polynucleotide sequence of apo-dystrophin-4 (SEQ ID NO: 2)

Figure 2



361

GGGAATGATTCCCAAATGGCAAAGAAAAGAGTGATGCTATCTATCTGCACCTTTGTA
-----+-----+-----+-----+-----+-----+-----+ 420
CCCTTACTAAAGGGTTACCGTTCTTGTCTCACTACGATAGATAGACGTGGAAAACAT
G N D F P N G K E T E * C Y L S A P F V

AAGTCTGTCTTCTTTCTCTTGTTCCAGGACACAATGTAGGAAGTCTTTCCACATG
421 -----+-----+-----+-----+-----+-----+-----+ 480
TTCAGACAGAAAGAAAGAGAAACAAAAGGTCTGTGTTACATCCTTCAGAAAAGGTGTAC
K S V F L S L C F P G H N V G S L F H M

GCAGATGATTGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA
481 -----+-----+-----+-----+-----+-----+-----+ 540
CGTCTACTAAACCGTCTCGCTACCTCAGGAATCATAGTCAGTACTGTCTACTTCTCCT
A D D L G R A M E S L V S V M T D E E G

GCAGAATAATGTTTACAACCTCCTGATTCCGCATGGTTTATAATATTACATAACA
541 -----+-----+-----+-----+-----+-----+-----+ 600
CGTCTTATTACAAAATGTTGAGGACTAAGGGCGTACCAAAATATTATAAGTATGTTGT
A E * M F Y N S * F P H G F Y N I H T T

AAGAGGATTAGACAGTAAGAGTTACAAGAAATAAATCTATATTGTGAAGGGTAGTG
601 -----+-----+-----+-----+-----+-----+-----+ 660
TTCTCCTAATCTGTCATTCTCAAATGTTCTTATTAGATATAAAACACTTCCCATCAC
K R I R Q * E F T R N K S I F L * R V V

GTATTATACTGTAGATTTCAGTAGTTCTAAGTCTGTTATTGTTAACATGGCAG
661 -----+-----+-----+-----+-----+-----+-----+ 720
CATATAATGACATCTAAAGTCATCAAAGATTCAAGACAATAACAAAACAATTGTTACCGTC
V L Y C R F Q * F L S L L L F C * Q W Q

Figure 2 (cont'd)



721 GTTTTACACGTCTATGCAATTGTACAAAAAGTTATAAGAAAACATGTAAAATCTG
721 -----+-----+-----+-----+-----+-----+ 780
CAAAATGTGCAGATACGTTAACATGTTTTCAATATTCTTTGATGTACATTTAGAAC
V L H V Y A I V Q K S Y K K T T C K I L

781 ATAGCTAAATAACTGCCATTCTTATATGGAACGCATTTGGGTTGTTAAAAATTAA
781 -----+-----+-----+-----+-----+-----+ 840
TATCGATTATTGAACGGTAAAGAAATATACCTTGCCTAAACCCAAACAATTAAAT
I A K * L A I S L Y G T H F G L F K N L

841 TAACAGTTATAAGAAAAGAATTATAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
841 -----+-----+-----+-----+-----+-----+ 900
ATTGTCAATATTCTTCTTAATATTCTTTCTTTATTGCCTTACCTGTTACCCAC
* Q L * R K N Y K G K R K * R N G Q V V

901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCAAAACCAAAGTGAGGTAGA
901 -----+-----+-----+-----+-----+-----+ 960
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTGTGGGTTTGGTTCACTCCATCT
K L * T Q V C T I I R N T P K P K * G R

961 AATAGCATGAGAAGCCGTGTTGATGTTAATTAAATT
961 -----+-----+-----+-----+ 996
TTATCGTACTCTTCGGCACAAACTACAATTAATTAA
N S M R S R V * C * L I

Figure 2 (cont'd)

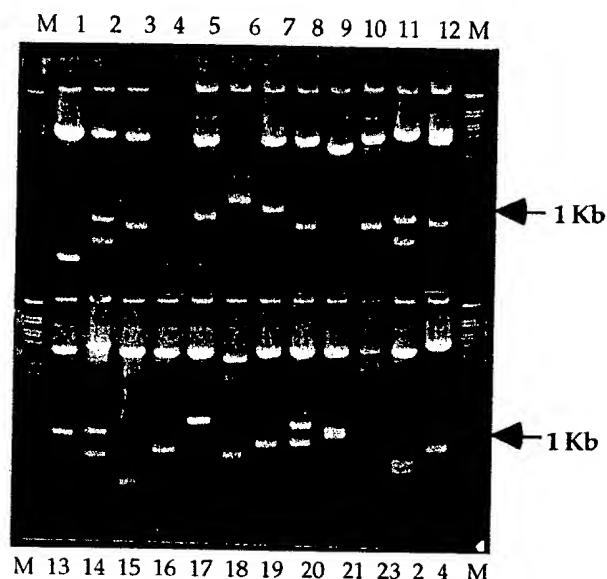


Figure 3A

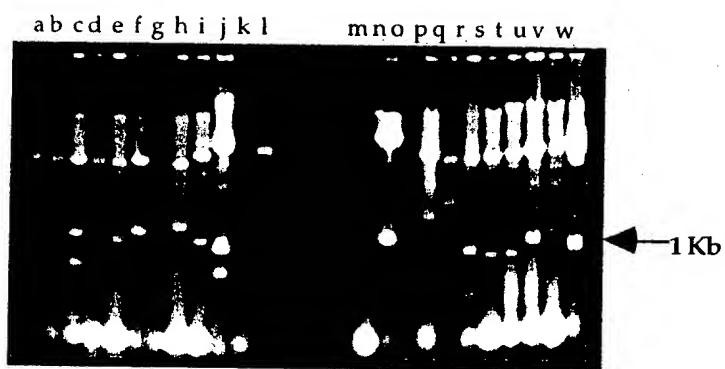


Figure 3B



SEP 18 2002

TECH CENTER 1600/2900

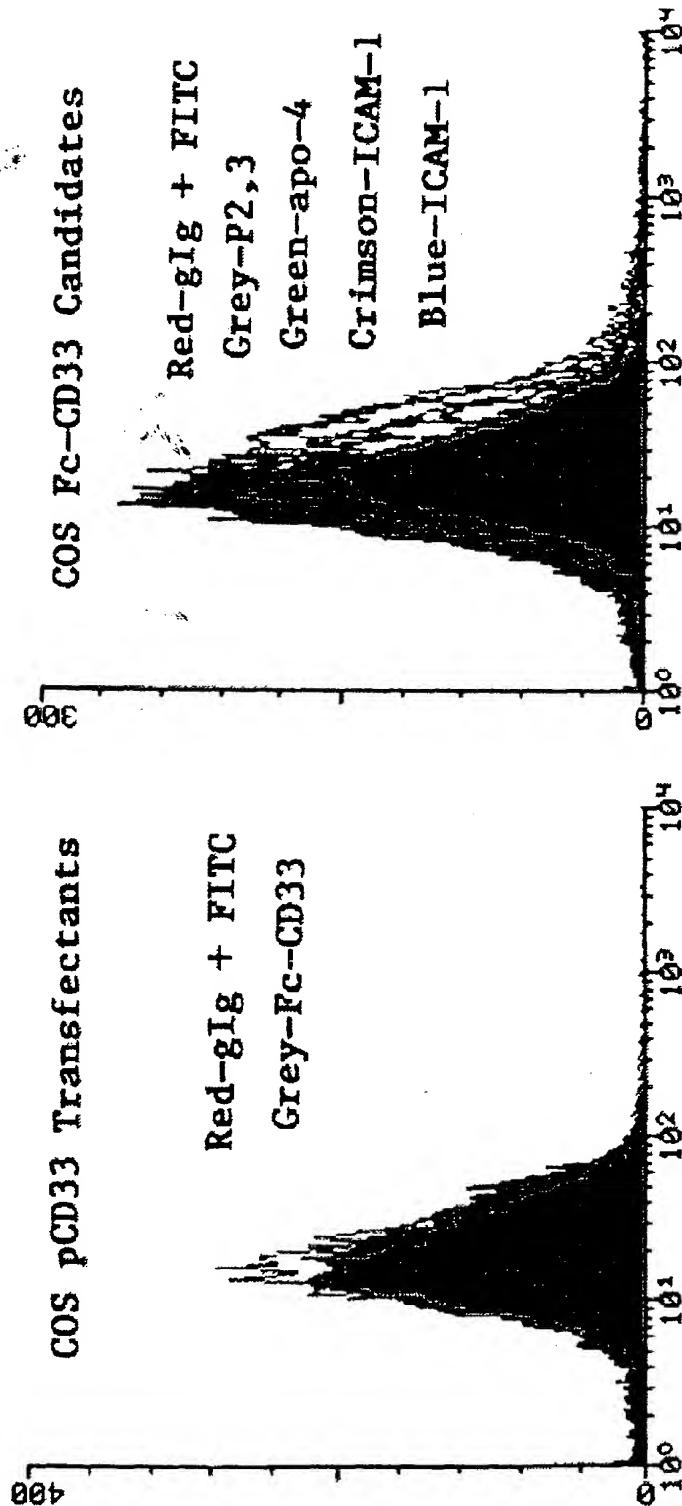


Figure 4A

Figure 4B

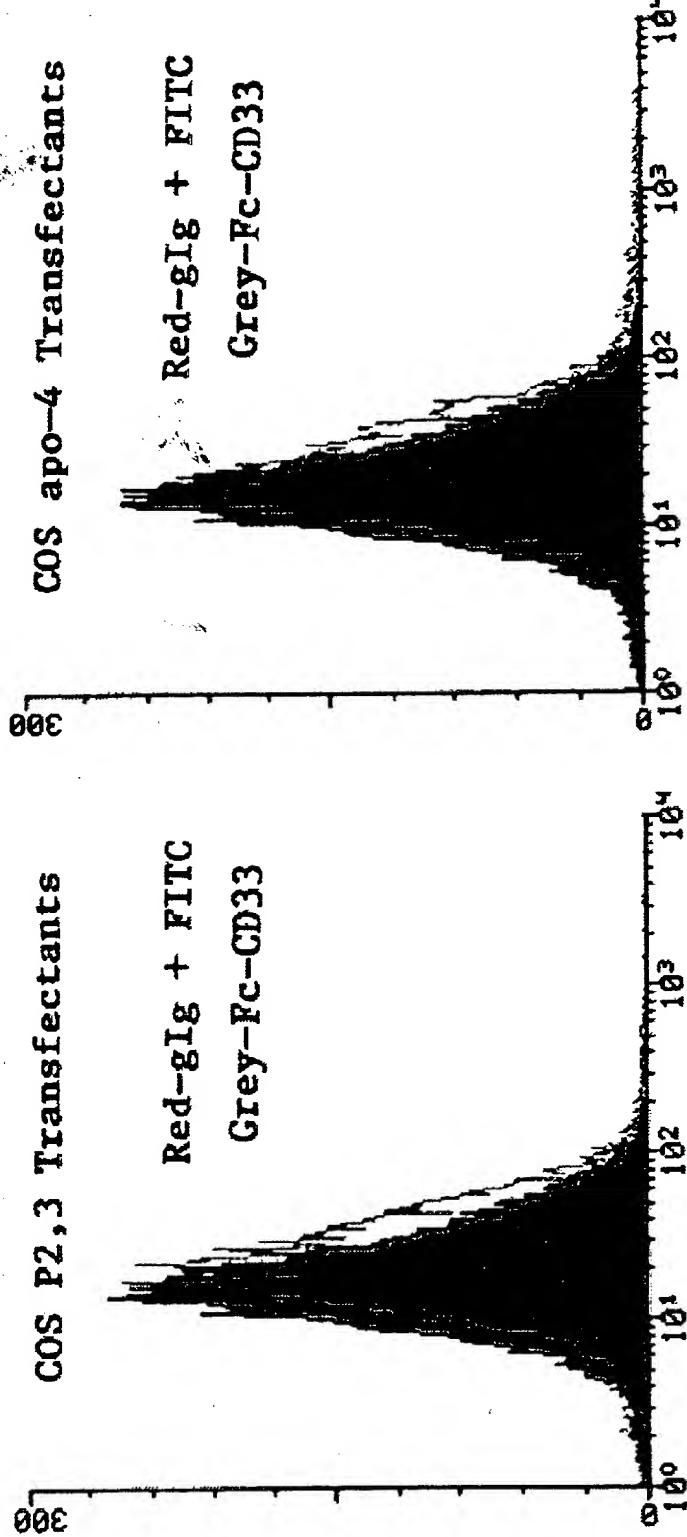


Figure 4C

Figure 4D

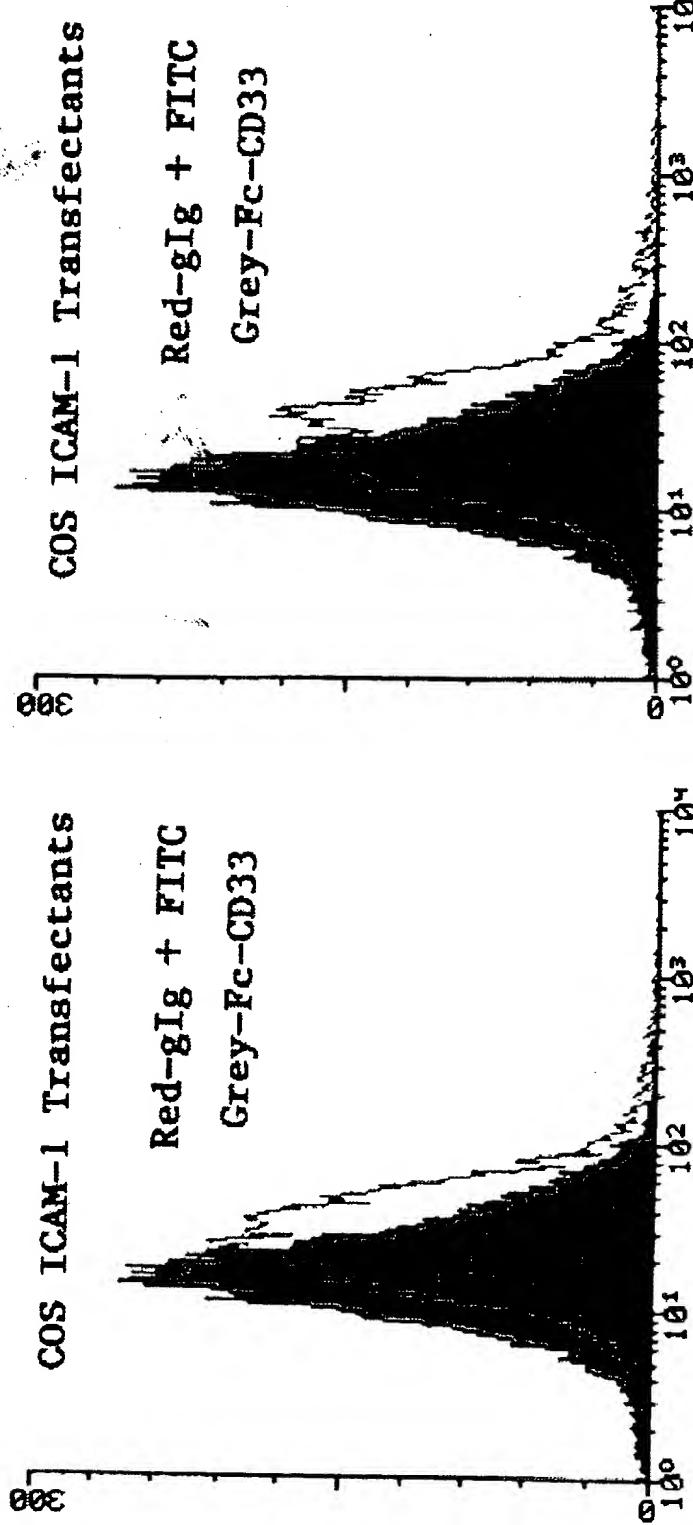


Figure 4E

Figure 4F

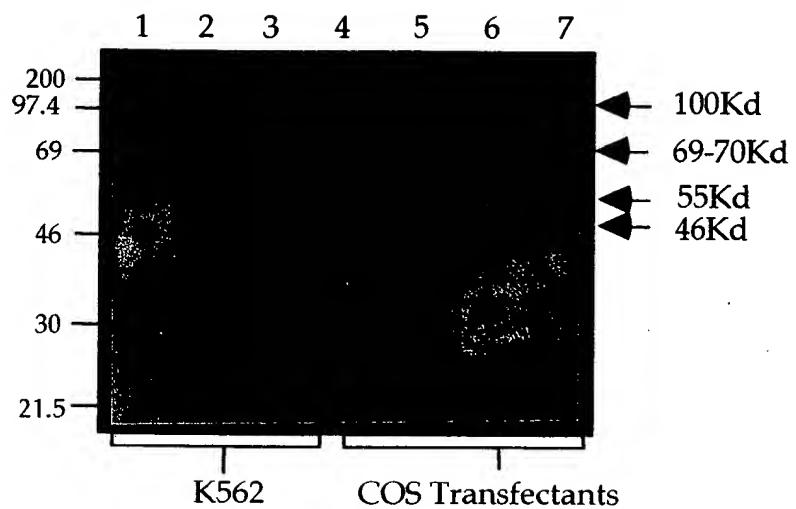


Figure 5



. . . T A G T T C C T A T T C A A T G T A T A G T G C A C C A A A G G T C A A T C A A G A G T T T A T T A T T
 -239 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 . . . A T C A A A G G A T A A G T T A C A T A T C A C G T G G T T C C A G T T A A G T T C T C A A A T A A T A A
 . * F P I Q C I V H Q R S I Q E F I I I

 A T T T C A A C C C A A G T A A A A G C A G A G A G A A A T A G C C A C C T C C A C C A T A G C C T C A G A A G C A
 -179 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 T A A A A G T T G G G T T C A T T T C G T C T C T C T T T A T C G G T G G A G G T G G T A T C G G A G T C T T C G T
 I F N P S K S R E K I A T S T I A S E A

 A G C C A A C A G C C T G A A A C A G C T T G A A A T G A A A G T T G G T G T G G C G G T G A T G G T G G C A G T G
 -119 -----+-----+-----+-----+-----+-----+-----+-----+
 T C G G T T G T C G G A C T T T G T C G A A A C T T T A C T T T C A A C C A C A C C G C C A C T A C C A C C G T C A C
 S Q Q P E T A L K * K V G V A V M V A V

 A T A A T G G T G A C C G A T G G T T G G G T G C T G G T G A T G G T A G T G G T A G T G T G A A G G T G G T G A T G
 -59 -----+-----+-----+-----+-----+-----+-----+-----+
 T A T T A C C A T G G C T A C C A A C C C A C G A C C A C T A C C A T C A C C A T C A A C A C T T C C A C C A C T A C
 I M V T D G W V L V M V V V V V V K V V M

 G T G G T T G A T T G A T A G T A A A A A A T G T C G T T A A T A C A A G T A G A G A G T A A G T A A T C A A T
 1 -----+-----+-----+-----+-----+-----+-----+-----+
 C A C C A A C T A A C T A C T A T T T T T A C A A G C A A T T A T G T T C A T C T C A T T C A T T A G T T A
 V V * L I V K K M F V N T S R E * V I N

 C A A T C A C T C A T A G C C A A G G T G G A A A A G A T G T A T C C C A T C A T G G A A T A T T C C T G T T C T G A T
 61 -----+-----+-----+-----+-----+-----+-----+-----+
 G T T A G T G A G T A T C G G T T C C A C C T T T C A T A G G G T A G T A C C T T A T A A G G A C A A G A C T A
 Q S L I A K V E K M Y P I M E Y S C S D

 A G A A A T C T T G T G C T T A T C T A T G G A A T T C T T T G A T A T A T A T T A C A T T G G G A A C C T G A A T
 121 -----+-----+-----+-----+-----+-----+-----+-----+
 T C T T T A G A A C A C G A A T A G A T A C C T T A A G A A A A C T A T A T A T A A A T G T A A C C C T T G G A C T T A
 R N L V L I Y G I L L I Y I Y I G N L N

 G T A G C T T G A C A T T T T C C A T G T A A A C A C C A G T A G C C T G A T C C A A C A T T A A G G C T G A T A C T A
 181 -----+-----+-----+-----+-----+-----+-----+-----+
 C A T C G A A C T G T A A A A G G T A C A T T T G T G G T C A T C G G A C T A G G G T G T A A T T C G A C T A T G A T
 V A * H F S M * T P V A * S N I K L I L

 A C A A A C A C G T G T A A T G G C T T C A T T A A T A A G G C T T G C T T C T G G A A A C T G G T G A A A
 241 -----+-----+-----+-----+-----+-----+-----+-----+
 T G T T T G T G C A C A T T A C C G A A G T A A T T A T T C G A A A C G A A G A A G G A C C T T G A C C A C T T
 T N N V * W L H * * G F A S S W K L V K

 A A T C A A A C C T T G T G T G A C A C C C T C G A T G C A G C T T C T G T G T G C T T C A C C C A G A A A T G
 301 -----+-----+-----+-----+-----+-----+-----+-----+
 T T A G T T G G A A C A A C A C A T G T G G G A G C T A C G T C G A A G A C A C A C A G A A G T G G G T C T T A C
 N Q T L L C T P S M Q L L C C L H P E M

Figure 6 (SEQ ID NO: 51)



361 GGGATGATTC CAAATGGCAAAGAACAGAGTGATGCTATCTATCTGCACCTTTGTA 420
 -----+-----+-----+-----+-----+
 CCCTTAAAGGTTTACCGTTCTTGTCTCACTACGATAGATAGACGTGGAAACAT
 G N D F P N G K E T E * C Y L S A P F V
 begin exon 79
 |
 421 AAGTCTGTCTTCTTCTCTTGTTCCAGGACACAATGTAGGAAGTCTTCCACATG 480
 -----+-----+-----+-----+-----+
 TTCAGACAGAAAGAAAGAGAAACAAAAGGTCTGTGTTACATCCTCAGAAAAGGTGTAC
 K S V F L S L C F P G H N V G S L F H M
 GCAGATGATTGGCAGAGCGATGGAGTCCTAGTATCAGTCATGACAGATGAAGAAGGA 540
 -----+-----+-----+-----+-----+
 CGTCTAAACCCGTCGCTACCTCAGGAATCATAGTCAGTACTGTCTACTTCTCCT
 A D D L G R A M E S L V S V M T D E E G
 GCAGAATAAAATGTTTACAACCTCTGATTCCCGATGGTTTATAATATTACATAACA 600
 -----+-----+-----+-----+-----+
 CGTCTTACAAAATGTTGAGGACTAAGGGCGTACCAAAATATTATAAGTATGTGT
 A E * M F Y N S * F P H G F Y N I H T T
 (----N----)
 601 AAGAGGATTAGACAGTAAGAGTTACAAGAAATAATCTATATTGTGAAGGGTAGTG 660
 -----+-----+-----+-----+-----+
 TTCTCCTAATCTGCATTCTCAAATGTTCTTATTAGATATAACACTTCCATCAC
 K R I R Q * E F T R N K S I F L * R V V
 GTATTATACTGTAGATTCAGTAGTTCTAAGTCTGTTATTGTTGTTAACATGGCAG 720
 -----+-----+-----+-----+-----+
 CATAATATGACATCTAAAGTCATCAAAGATTCAAGACAATAACAAAACAATTGTTACCGTC
 V L Y C R F Q * F L S L L L F C * Q W Q
 GTTTACACGTCTATGCAATTGTACAAAAAGTTATAAGAAAATACATGTTAACATGGCAG 780
 -----+-----+-----+-----+-----+
 CAAATGTGCAGATACGTTAACATGTTCTCAATATTCTTGTACATTAGAAC
 V L H V Y A I V Q K S Y K K T T C K I L
 ATAGCTAAATAACTGCCATTCTTATGGAACGCATTTGGGTTGTTAAAAATTAA 840
 -----+-----+-----+-----+-----+
 TATCGATTATTGAACGGTAAAGAAATATACCTTGCCTAAACCAACAAATTAAAT
 I A K * L A I S L Y G T H F G L F K N L
 inversion start site
 |
 841 TAACAGTTATAAGAAATTATAAGAAAAAGAAAATACGCAATGGACAAGTGGTG 900
 -----+-----+-----+-----+-----+
 ATTGTCAATATTCTTCTTAAATATTCTTCTTATTGCGTTACCTGTTACCCAC
 * Q L * R K N Y K G K R K * R N G Q V V

Figure 6 (cont'd)



901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCCAAAACCAAAGTGAGGTAGA
-----+-----+-----+-----+-----+-----+-----+-----+

TTCGACACTTGAGTCCACACGTGTTAATAGTCCTGTGGGGTTTGGTTTCACTCCATCT
K L * T Q V C T I I R N T P K P K * G R

AATAGCATGAGAAGCCGTGGATGTTAATTAATT

961 -----+-----+-----+----- 996

TTATCGTACTCTTCGGCACAAACTACAATTAATT
N S M R S R V * C * L I

Figure 6 (cont'd)

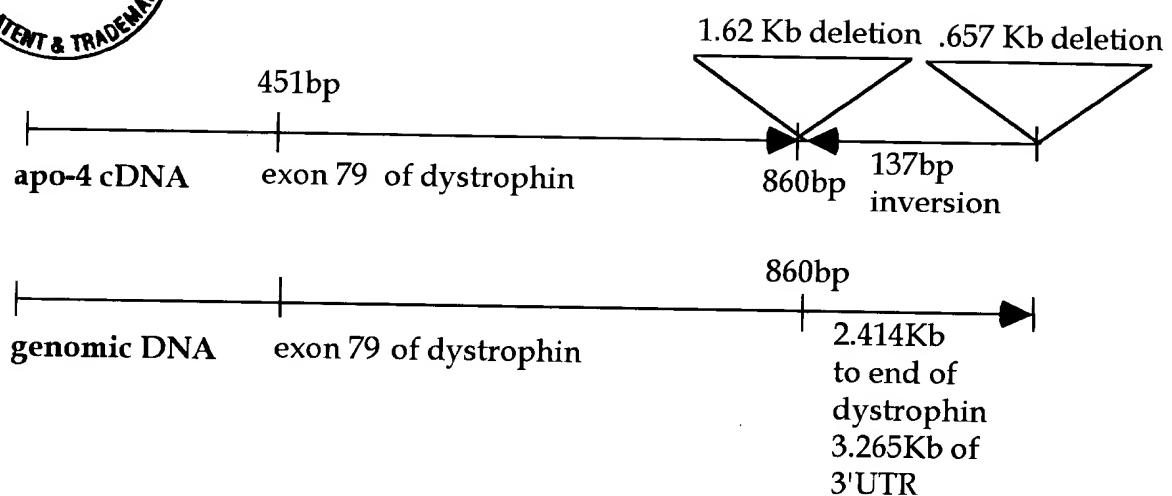
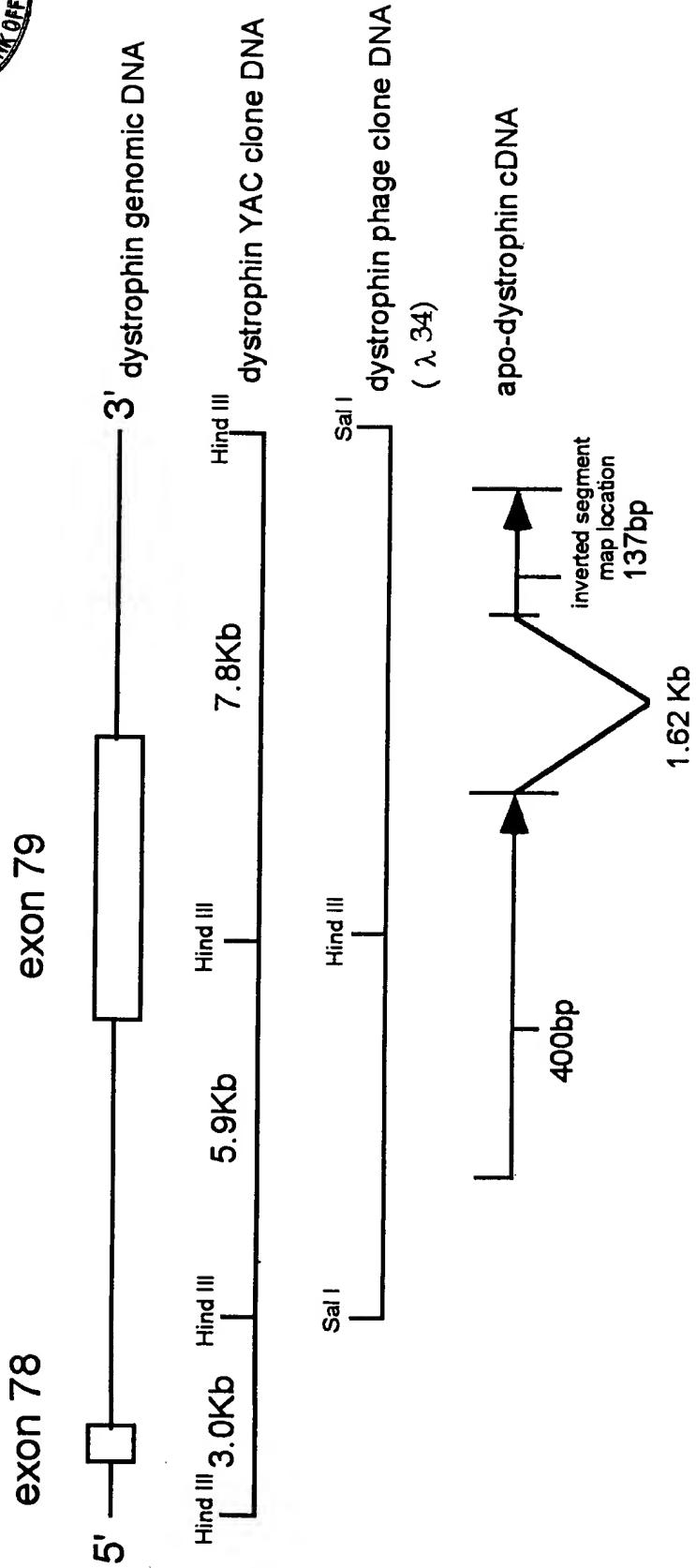


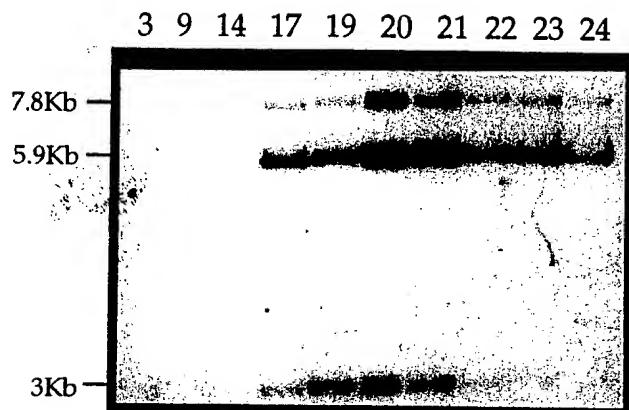
Figure 7



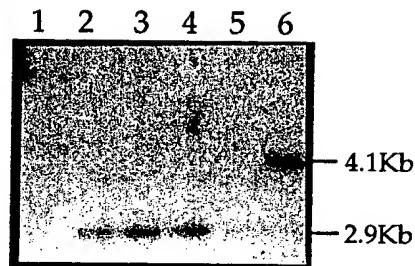
*cDNA map is not precisely drawn to scale

Figure 8

OIR E JC 135
SEP 16 2002
PATENT & TRADEMARK OFFICE



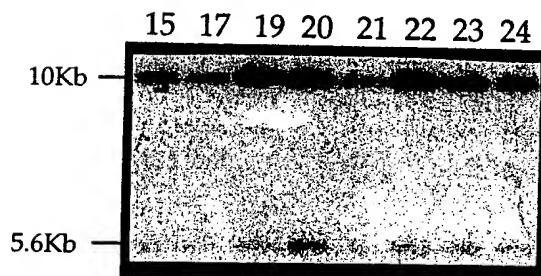
A. Human



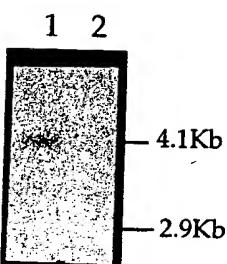
B. Mouse

Figure 9A

Figure 9B



A. Human



B. Mouse

Figure 10A

Figure 10B



RECEIVED

SEP 18 2002

TECH CENTER 1600/2900

	1		50
ID NO 54)	Mgen1073	
(SEQ ID NO 55)	Hapo1234	ctagttcctt attcaatgtatgtcacca aaggtaattt caagat	
Consensus		-----	-----
	51		100
Mgen1073	
Hapo1234	ttattattat tttcaaccca	agtaaaaagca gagagaaaaat agccacatcc	
Consensus	-----	-----	-----
	101		
Mgen1073	ttcACAGgCT tAAgCAGCca gtAAATGAcA	
Hapo1234	accatagcct cagaagcaag ccaACAGCCT gAAaCAGCtt tgAAATGA <u>a</u> A		
Consensus	-----	-----ACAG-CT -AA-CAGC-- AAATGA-A	
	151		200
Mgen1073	AtT.....T AtgtGgtAgt cAgGtcactG	
Hapo1234	AgTtgggtgtg	gcccgtatgg tggcagtgtat AatgGtgAcc gAtGgtggG	
Consensus	A-T-----	-----T A---G--A-- -A-G-----G	
	201		
Mgen1073	TGCTGGTaAT GGTgaTctTA	GcaGgcAgAG aaGGTGgTaG TGaTTGATA	
Hapo1234	TGCTGGTgAT GGTgTggTA	GttGtgA.AG gtGGTGaTgG TGgTTGATT	
Consensus	TGCTGGT-AT GGT--T--TA	G--G--A-AG --GGTG-T-G TG-TTGAT-	
	251		300
Mgen1073	GtaAaagtgt AgActaTaCa	acAgaAtAAa TAcAagtatA GTAA.....	
Hapo1234	GatAgtaaaa AaAa <u>TgTtCg</u>	ttAatAcAAg TAgAgatA GTAAAtcaatc	
Consensus	G--A----- A-A-T-T-C-	--A--A-AA- TA-A----A GTAA-----	
	301	M1	
Mgen1073atc caaCAAaGTG	tgAAAGgTGT gTgCCATtAc acAtctTTCT	
Hapo1234	aatcactcat agcCAAgtGTG	gaAAAGaTGT aTcCCATcAt ggAataTTCc	
Consensus	-----	---CAA-GTG --AAAG-TGT -T-CCAT-A- --A---TTC-	
	351	M2	
Mgen1073	cG..... GtgATAagag	cCTTgTCTAT GaAgTTC... TGAgATgTgT	
Hapo1234	tGttctgata GaaATcttgc	gCTTatCTAT GgAaTTCTtt TGAtATAatAT	
Consensus	-G----- G--AT-----	-CTT-TCTAT G-A-TTC--- TGA-AT-T-T	
	401		450
Mgen1073	TaggAagatG AAtCatcAat	TtaCaT.... TTcTcCCcat cAAAtgaCAC	
Hapo1234	TtacAttggG AAcCtgaAtg	TagCtTgaca TTtTtCCatg tAAAacacCAg	
Consensus	T---A----G AA-C---A--	T--C-T---- TT-T-CC--- -AAA---CA-	
	451		500
Mgen1073	cAtgCTGATC CAgtATTAAG	CTaATACTAA C.....ACca tgcAatGCTT	
Hapo1234	tAgcCTGATC CAacATTAAG	CTgATACTAA CaaacaACgt gtaAtgGCTT	
Consensus	-A--CTGATC CA--ATTAAG	CT-ATACTAA C----AC-- ---A--GCTT	
	501		550
Mgen1073	CATTAACAcAAG GaTTTGCTTC	TTgCTaGAAA tgGGT. .AAA AaCggACtgT	
Hapo1234	CATTAAtAAG GcTTTGCTTC	TTcCTgGAAA ctGGTgaAAA AtCaaACctT	
Consensus	CATTAA-AAG G-TTTGCTTC	TT-CT-GAAA --GGT---AAA A-C--AC--T	

Figure 11



-70 bp from 5' end of apo-4

|

Inr = GCCC TCAT TCTG GAGAC (SEQ ID NO: 59)

apo-4 = GCGG TGAT GGTG GCAGT (SEQ ID NO: 60)

- 48% perfect homology with Inr

71% match on type of base

(purine vs. pyrimidine)

Figure 12B



	551				
Mgen1073	GgTcTGTAtA	CCTTCaATGC	AGCTTaTGTG	TTGTCTTtC	C..tgAAatG
Hapo1234	GtTgTGTAcA	CCcTCgATGC	AGCTTcTGTG	TTGTCTTcaC	CcagaAAAtgG
Consensus	G-T-TGTA-A	CC-TC-ATGC	AGCTT-TGTG	TTGTCTT--C	C----AA--G
	601				
Mgen1073	GtAATGAcTc	CCaAtAgtGg	cAAccAgggG	tacaATaCT.TGCA
Hapo1234	GgAATGAtTt	CCcAaAtgGc	aAgaAacaG	agtgATgCTa	tctatcTGCA
Consensus	G-AATGA-T-	CC-A-A--G-	-AA--A---G	----AT-CT-	-----TGCA
	651				exon79 700
Mgen1073	CacTTTGTAA	A....cTCTT	TCTTTCTCTT	TGTTTTCCAG	GACACAATGT
Hapo1234	CctTTTGTAA	AgtctgTCTT	TCTTTCTCTT	TGTTTTCCAG	GACACAATGT
Consensus	C--TTTGTAA	A----TCTT	TCTTTCTCTT	TGTTTTCCAG	GACACAATGT
	701				750
Mgen1073	AGGAAGGcCTT	TTCCACATGG	CAGATGATT	GGGCAGAGCG	ATGGAGTCCT
Hapo1234	AGGAAGGtCTT	TTCCACATGG	CAGATGATT	GGGCAGAGCG	ATGGAGTCCT
Consensus	AGGAAGG-CTT	TTCCACATGG	CAGATGATT	GGGCAGAGCG	ATGGAGTCCT
	751				800
Mgen1073	TAGTtTCAGT	CATGACAGAT	GAAGAAGGAG	CAGAATAAAAT	GTTTTACAAC
Hapo1234	TAGTatTCAGT	CATGACAGAT	GAAGAAGGAG	CAGAATAAAAT	GTTTTACAAC
Consensus	TAGT-TCAGT	CATGACAGAT	GAAGAAGGAG	CAGAATAAAAT	GTTTTACAAC
	801				850
Mgen1073	TCCTGATTCC	CGCATGGTTT	TTATAATATT	CgTACAACAA	AGAGGATTAG
Hapo1234	TCCTGATTCC	CGCATGGTTT	TTATAATATT	CaTACAACAA	AGAGGATTAG
Consensus	TCCTGATTCC	CGCATGGTTT	TTATAATATT	C-TACAACAA	AGAGGATTAG
	851				900
Mgen1073	ACAGTAAGAG	TTTACAAGAA	AtaAAATCTA	TATTTTGTG	AAGGGTAGTG
Hapo1234	ACAGTAAGAG	TTTACAAGAA	AT.AAATCTA	TATTTTGTG	AAGGGTAGTG
Consensus	ACAGTAAGAG	TTTACAAGAA	AT-AAATCTA	TATTTTGTG	AAGGGTAGTG
	901				950
Mgen1073	GTAcTATACT	GTAGATTCA	GTAGTTCTA	AGTCTGTTAT	TGTTTGTGA
Hapo1234	GTAtTATACT	GTAGATTCA	GTAGTTCTA	AGTCTGTTAT	TGTTTGTGA
Consensus	GTA-TATACT	GTAGATTCA	GTAGTTCTA	AGTCTGTTAT	TGTTTGTGA
	951				1000
Mgen1073	ACAATGGCAG	GTTCACACG	TCTATGCAAT	TGTACAAAAA	AGTTAAAGA
Hapo1234	ACAATGGCAG	GTTCACACG	TCTATGCAAT	TGTACAAAAA	AGTTAtAAGA
Consensus	ACAATGGCAG	GTTCACACG	TCTATGCAAT	TGTACAAAAA	AGTTA-AAGA
	1001				1050
Mgen1073	AA...ACATG	TAAAATCTTG	ATAGCTAAAT	AACTGCCAT	TTCTTTATAT
Hapo1234	AAactACATG	TAAAATCTTG	ATAGCTAAAT	AACTGCCAT	TTCTTTATAT
Consensus	AA--ACATG	TAAAATCTTG	ATAGCTAAAT	AACTGCCAT	TTCTTTATAT

Figure 11 (cont'd)



		begin inversion@1100
	1051	1100
Mgen1073	GGAACGCATT TTGGGTTGTT TAAAAATTAA TAACAGTTAT AAAGAAAGAT	
Hapo1234	GGAACGCATT TTGGGTTGTT TAAAAATTAA TAACAGTTAT AAAGAAAGAa	
Consensus	GGAACGCATT TTGGGTTGTT TAAAAATTAA TAACAGTTAT AAAGAAAGA-	
	1101	1150
Mgen1073	TgtAAActaA Agtgtgcttt AtAAAAaAAg ttgtTtataA AaaccctAa	
Hapo1234	TtatAAaggA A.....aa AgAAAAAtAAc gcaaTggacA AgtgtgtaAg	
Consensus	T---AA---A A----- A-AAAA-AA- ----T---A A-----A-	
	1151	1200
Mgen1073	acaaacACaC AcGcacaCAC AcacAcacac AcacaCaCAC AcaCACAcTG	
Hapo1234	ctgtgaACTC AgGtgtgCAC AattAtcagg AacacCcCAa AacCAaAgTG	
Consensus	-----AC-C A-G---CAC A---A---- A---C-CA- A--CA-A-TG	
	1201	1243
Mgen1073	AGGcAGcAca ttgtTttGcA ttacTtTagc gTGTatcaTA t..	
Hapo1234	AGGtAGaAat agcaTgaGaA gccgTgTttg aTGTtaatTA att	
Consensus	AGG-AG-A-- ----T--G-A ----T-T--- -TGT----TA ---	

Figure 11 (cont'd)

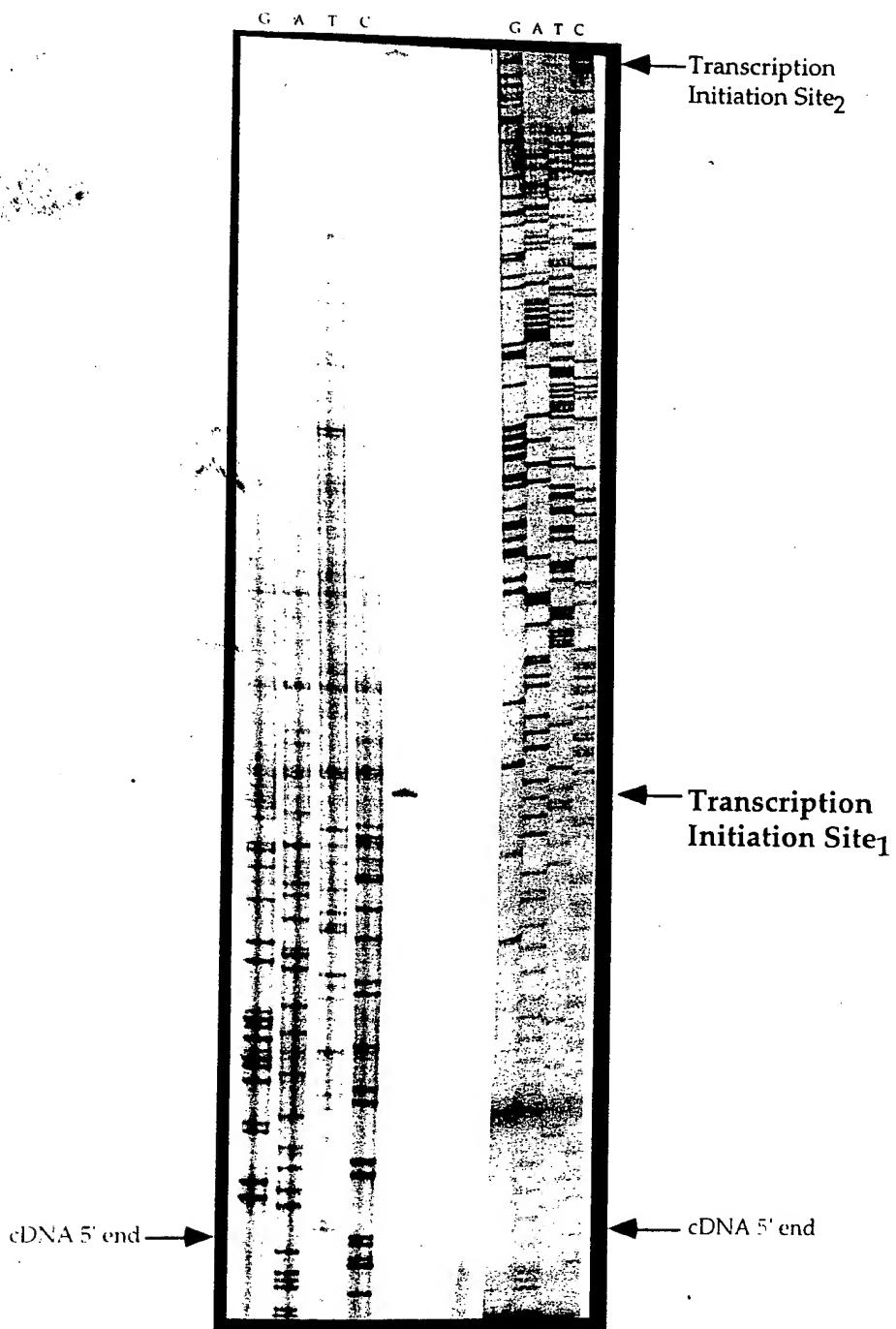


Figure 12A

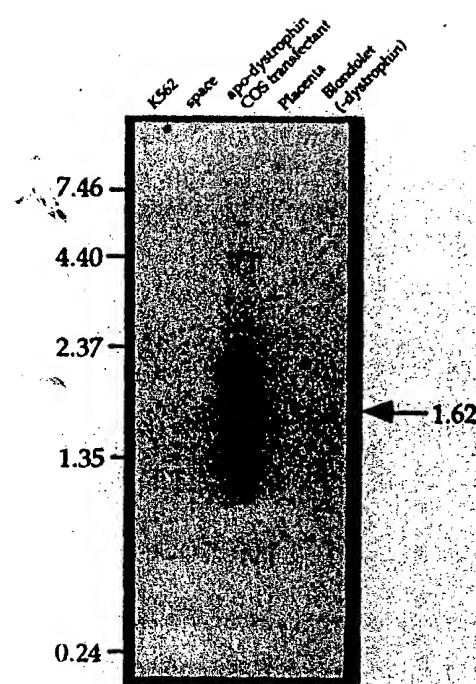


Figure 13

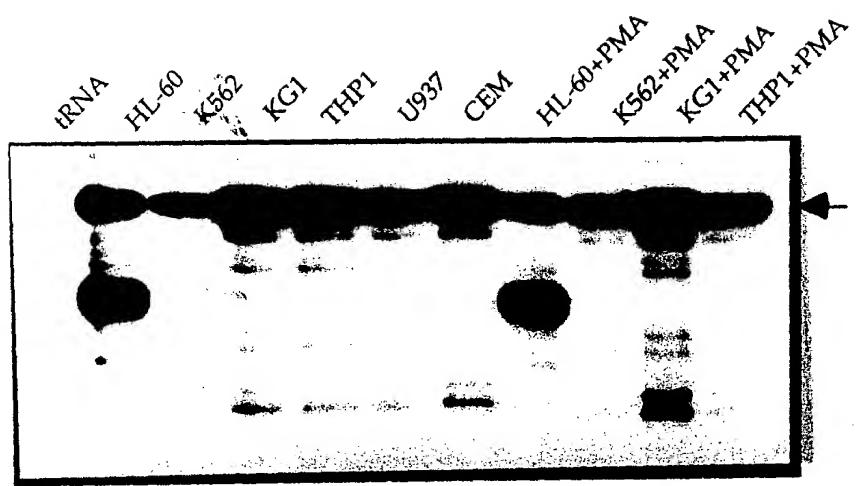


Figure 14

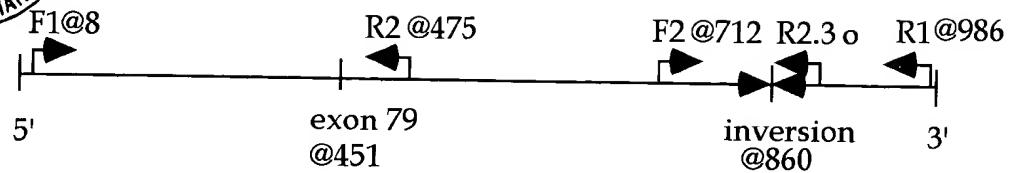


Figure 15

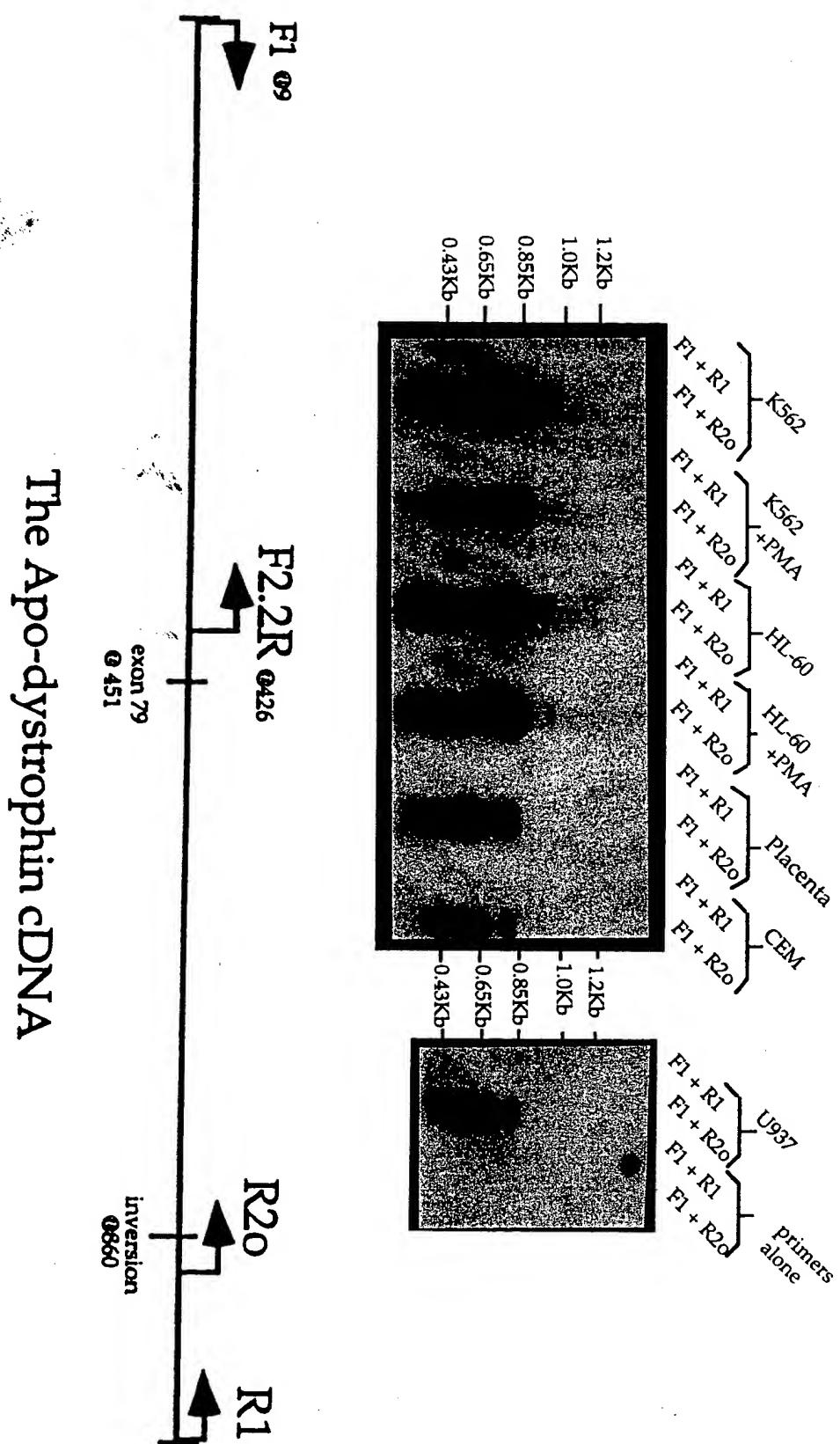


Figure 16

The Apo-dystrophin cDNA

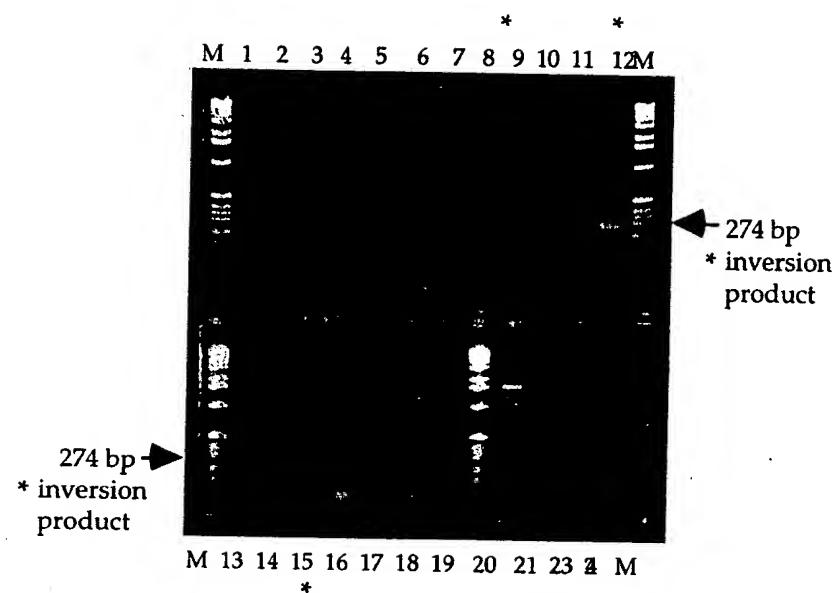
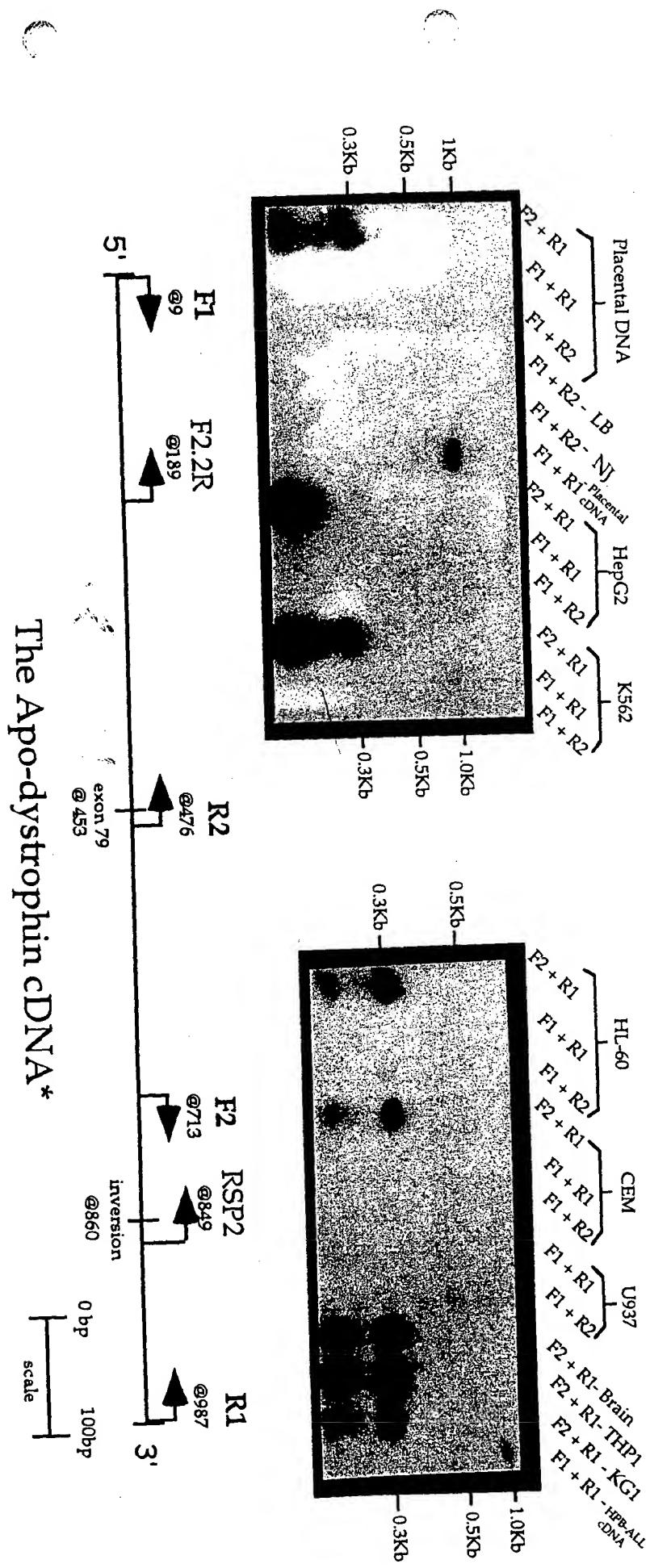


Figure 17A



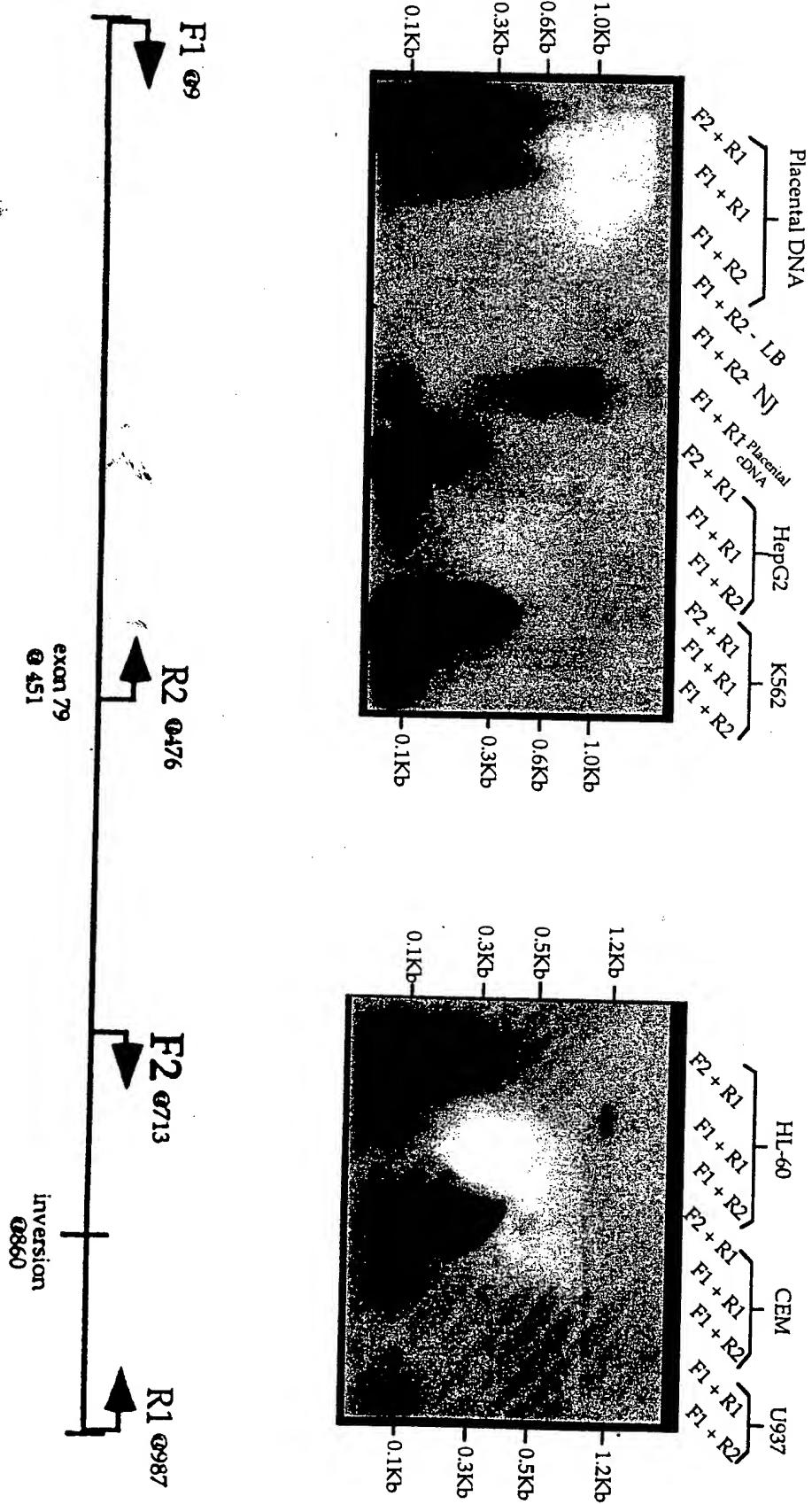
*RT-PCR primers (**bold**) and Southern Blotting probes are indicated by the arrows.

Figure 17B



The Apo-dystrophin cDNA

Figure 17C





RECEIVED

SEP 18 2002

TECH CENTER 1600/2900

A.

12/23bp spacer
CACAGTG-----ACAAAAAACC
 heptamer nonamer

Figure 18A

B.

inversion breakpoint₁

	11640	11650	11660	11670	11680	
	*	*	*	*	*	
dystrophin	T TTATAACAGT	<u>TATAAAGAAA</u>	GA^TTGTAAAC	<u>TAAAGTGTGC</u>		
	A AATATTGTCA	ATATTCTTT	CT^AACATTG	ATTTCACACG		
apo-4 cDNA	840	850		870		a
[138]	T TTATAACAGT	<u>TATAAAGAAA</u>	GA^TTaTAAAg	gAAAaaGaaa>		
	^ ^ ^ ^ ^ ^ ^	^ ^ ^ ^ ^ ^ ^	^ ^ v ^ ^ ^ v	v ^ ^ v v ^ v v v		
dystrophin	T TTATAACAGT	<u>TATAAAGAAA</u>	GA^TTGTAAAC	TAAAGTGTGC		

	11690	11700	11710	11720	11730	
	*	*	*	*	*	
dystrophin	<u>TTTATAAAAA</u>	<u>AAAGTTGTTT</u>	<u>ATAAAAACCC</u>	<u>CTAAAAACAA</u>	<u>AACAAACACA</u>	
	AAATATTTTT	TTTCAACAAA	TATTTTGCGG	GATTTTGTT	TTGTTTGTT	
apo-4 cDNA	880	890	900	910	920	930
[138]	aTaAaAtggA	cAAAGTgGTga	ATgtgAACtC	agggtgtgCAc	AAttAtCAGg>	
	v^v^v^v^v^v^	v^v^v^v^v^v^	v^v^v^v^v^v^	v^v^v^v^v^v^v^	v^v^v^v^v^v^v^	
dystrophin	TTTATAAAAA	AAAGTTGTTT	ATAAAAACCC	CTAAAAACAA	AACAAACACA	

	11740	11750	
	*	*	
dystrophin	<u>CACACACACA</u>	<u>CATACACACA</u>	
	GTGTGTGTGT	GTATGTGTGT	
apo-4 cDNA	940	950	
[138]	aACAC-CcCA	-AaAC-CAaA>	
	v^v^v^v^	v^v^v^v^	
dystrophin	CACACACACA	CATACACACA	

Figure 18B (SEQ ID NO: 57)



SEP 16 2002

PATENT & TRADEMARK OFFICE

inversion breakpoint₂

13130	13140	13150	13160	13170
*	*	*	*	*

dystrophin AATTAGCTTT TGGAGAGTGG GTTTTGTCCA TTATTAATAAA TTAATTAATT
 TTAATCGAAA ACCTCTCACC CAAACAGGT AATAATTATT AATTAATTAA

apo-4

990

<AATTAATT
~~~~~

dystrophin

AATTAATT

|       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 13180 | 13190 | 13200 | 13210 | 13220 |
| *     | *     | *     | *     | *     |

dystrophin    AACATCAAAC ACGGCTTCTC ATGCTATTTC TACCTCACTT TGGTTTGGG  
TTGTAGTTTG TGCCGAAGAG TACGATAAAG ATGGAGTGAA ACCAAAACCC

apo-4

980            970            960            950            940  
<AACATCAAAC ACGGCTTCTC ATGCTATTTC TACCTCACTT TGGTTTGGG  
~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~

dystrophin AACATCAAAC ACGGCTTCTC ATGCTATTTC TACCTCACTT TGGTTTGGG

| | | | | |
|-------|-------|-------|-------|-------|
| 13230 | 13240 | 13250 | 13260 | 13270 |
| * | * | * | * | * |

dystrophin GTGTTCTGA TAATTGTGCA CACCTGAGTT CACAGCTTCA CCACTTGTCC
 CACAAGGACT ATTAACACGT GTGGACTCAA GTGTCGAAGT GGTGAACAGG

apo-4

930 920 910 900 890
<GTGTTCTGA TAATTGTGCA CACCTGAGTT CACAGCTTCA CCACTTGTCC
~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~

dystrophin    GTGTTCTGA TAATTGTGCA CACCTGAGTT CACAGCTTCA CCACTTGTCC

|       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 13280 | 13290 | 13300 | 13310 | 13320 |
| *     | *     | *     | *     | *     |

dystrophin    ATTGCGTTAT TTTCTTTTTC CTTTATAATT CTTTCTTTTT CCTTCATAAT  
TAACGCAATA AAAGAAAAAG GAAATATTAA GAAAGAAAAA GGAAGTATTA

**inversion breakpoint<sub>3</sub>**

|     |     |     |     |     |
|-----|-----|-----|-----|-----|
| 880 | 870 | 860 | 850 | 840 |
|-----|-----|-----|-----|-----|

apo-4    <ATTGCGTTAT TTTCTTTTTC CTTTATAATT CTTTCTTTTaT aacTgtTAta  
~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ^v^ vvv^vv^vv

dystrophin ATTGCGTTAT TTTCTTTTTC CTTTATAATT CTTTCTTTTT CCTTCATAAT

Figure 18C (SEQ ID NO: 58)

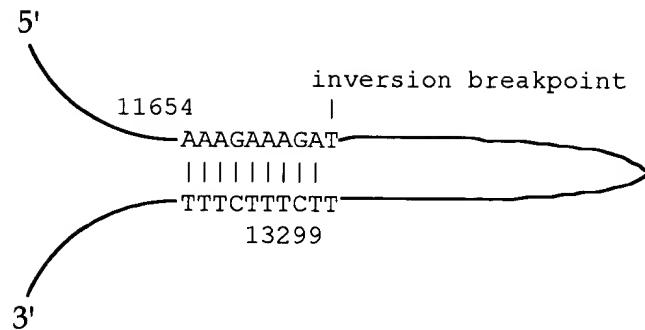


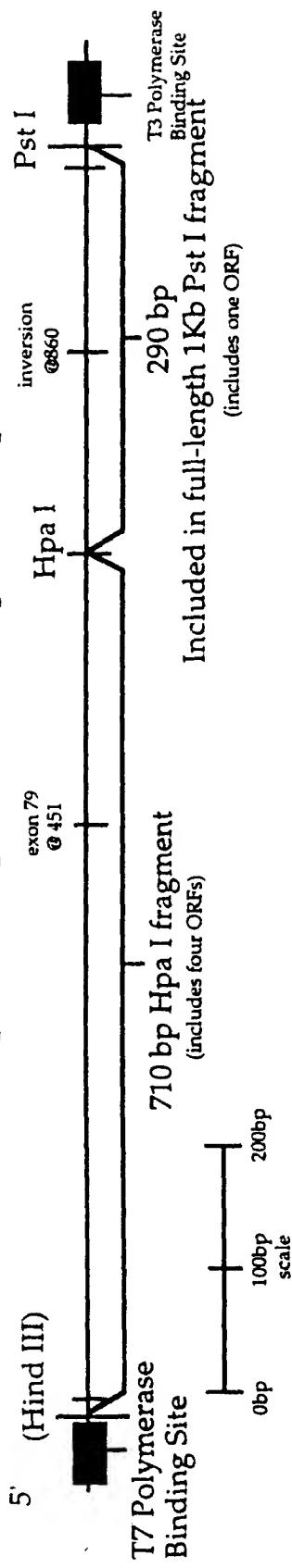
Figure 18D (Sequence included in SEQ ID NO: 57)

inversion @ 860
841 TAACAGTTATAAAGAAAGAAT TATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
ATTGTCAATTTTCTTCTTAATTTCCCTTTCTTTATTGCGTACCTGTTACCCAC 900

Figure 19 (Sequence included in SEQ ID NO: 2)



**Description of *In Vitro* Transcription and Translation
of the Apo-dystrophin cDNA in pBluescript SK+**



Linearize plasmid with either HpaI (truncated) or PstI (full length). Gene Clean and incubate with T7 polymerase and dNTPs to produce RNA in vitro.

↓

Incubate RNA with Wheat Germ Extract or Rabbit Reticulocyte Lysates to produce *in vitro* translation

↓

Separate translation products by SDS-PAGE. Fix, Amplify and Dry Gel. Perform Autoradiography

Figure 20

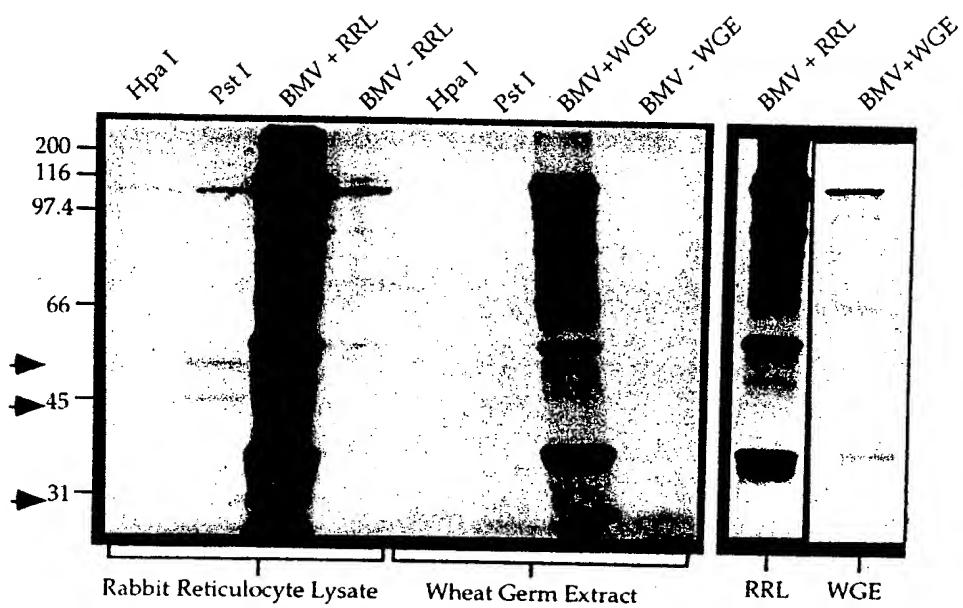


Figure 20A

Figure 20B

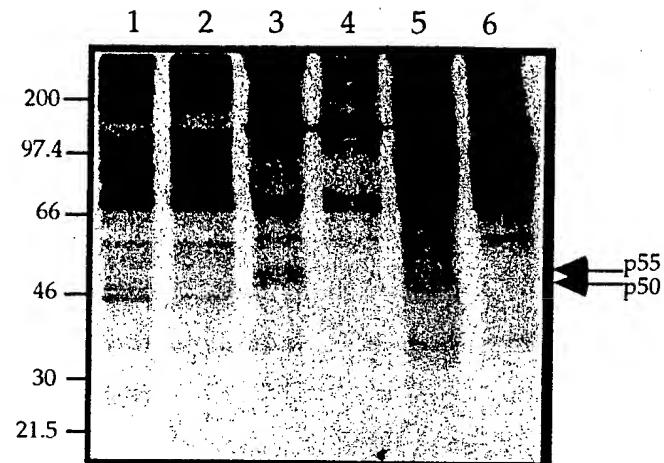


Figure 21

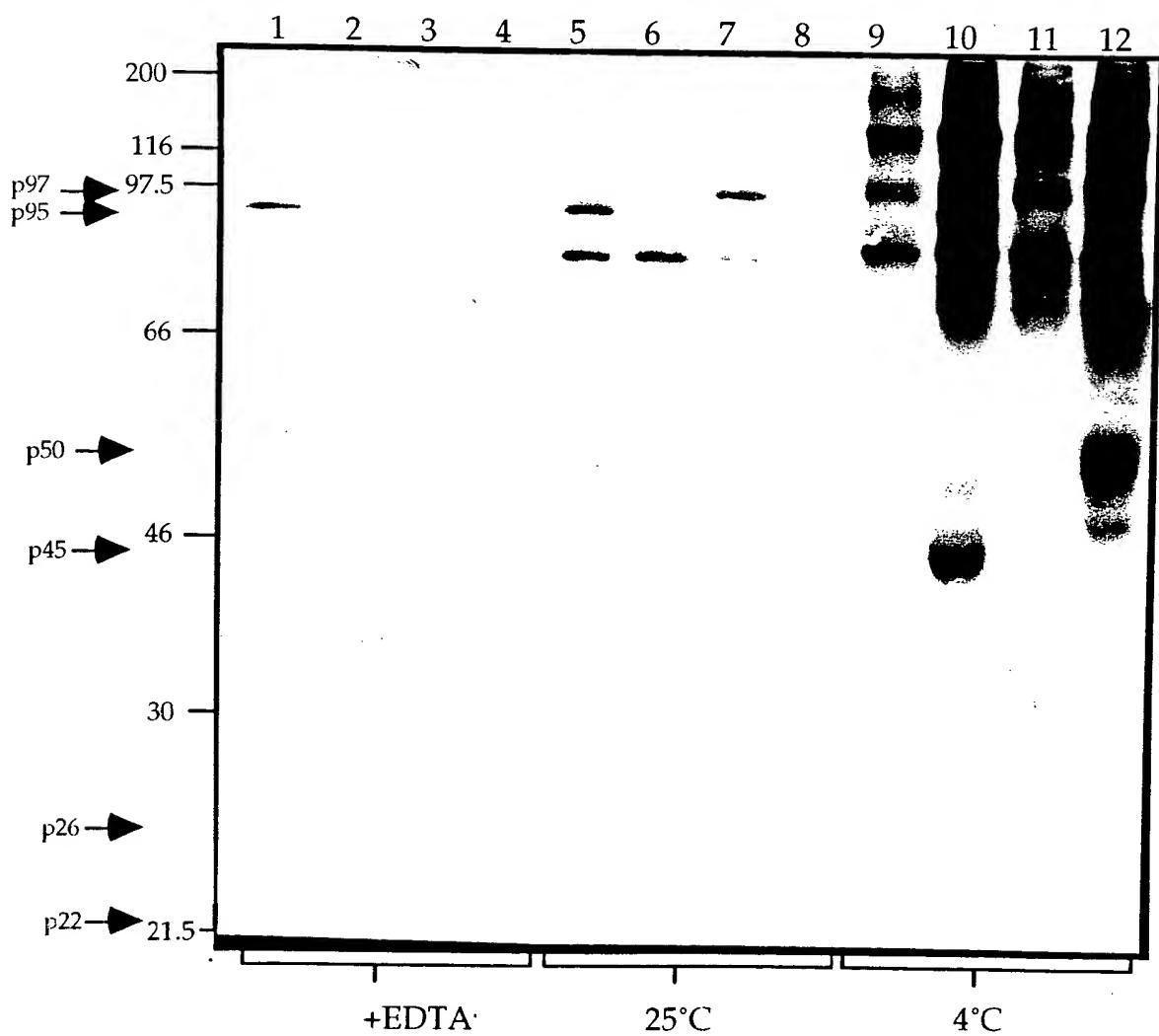


Figure 22

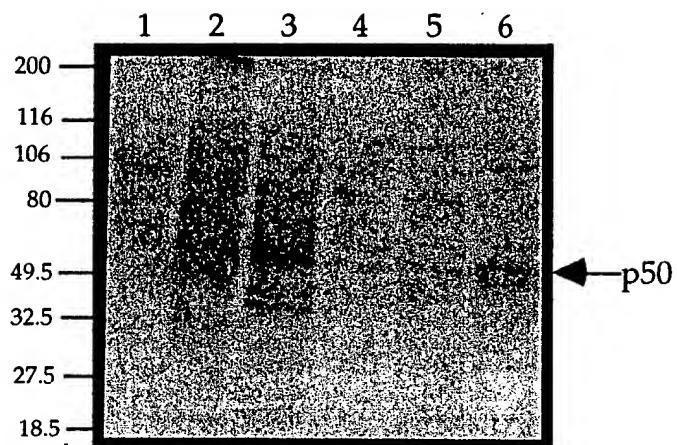


Figure 23

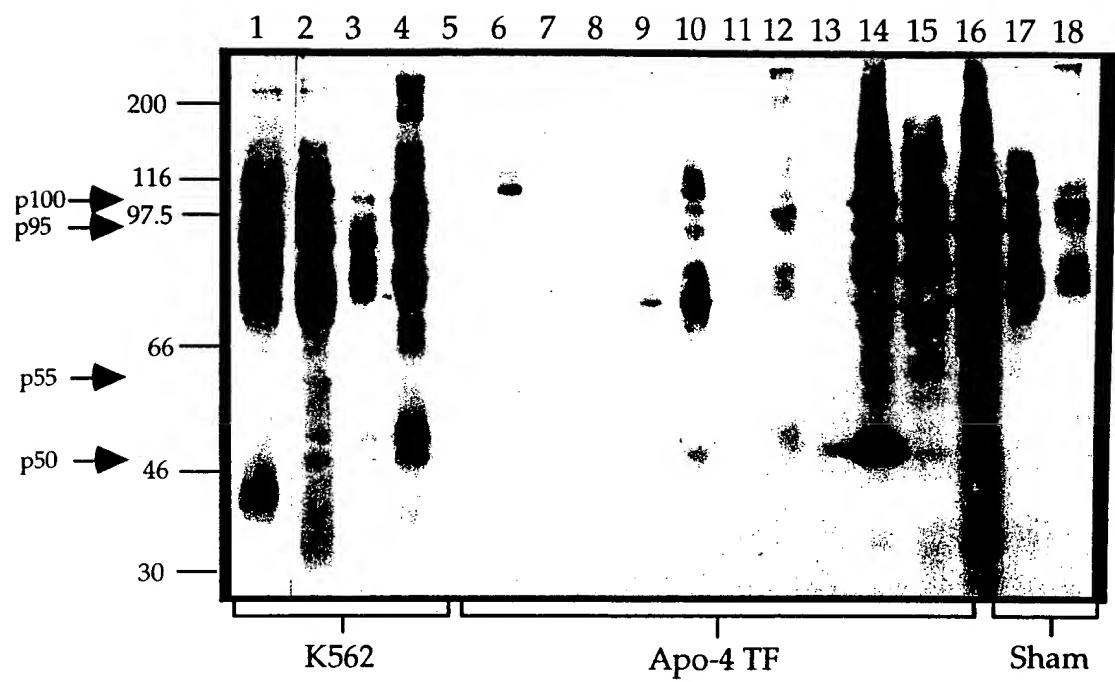


Figure 24

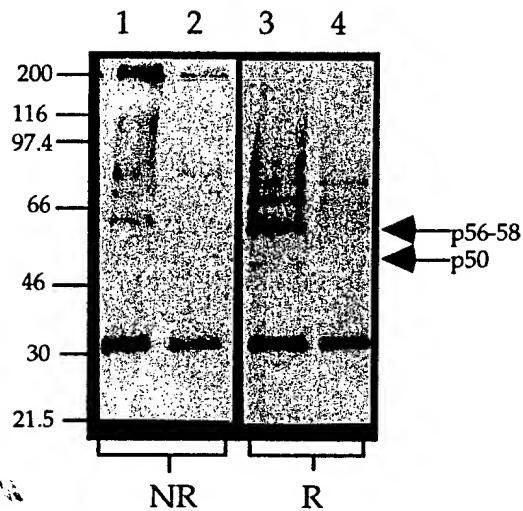


Figure 25A

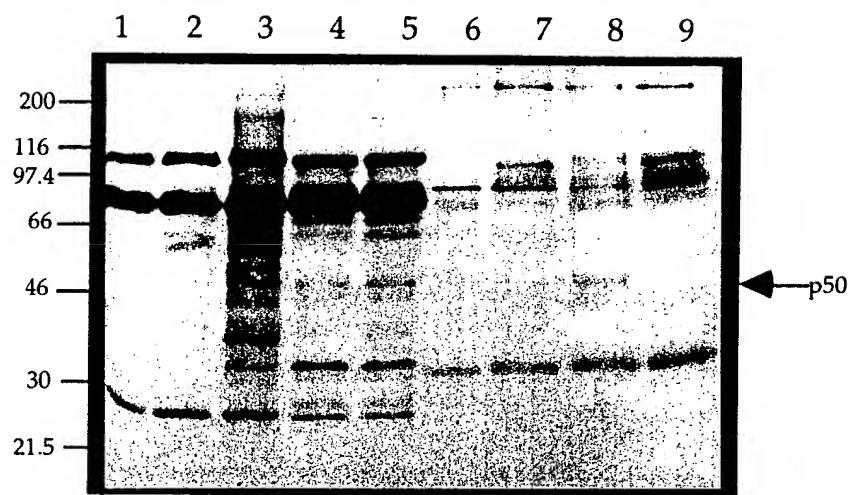


Figure 25B

H2 starting at second methionine - 321 bp, predicted weight = 17.4Kd + 1 N-glycosylation site + 20.4 Kd.



Figure 26A

Splice sites for peptide

MYPIMEYSCSDRNLVLIYGILLIYIYIGNLNMKKEQNKCFTTPDSRMVFIIFI
QQRGLDSKSLQEINLYFCEGFYTSMLYKKVIRKLHKITQWTRTPQNQSEV
EIA (SEQ ID NO: 61)

Figure 26B

| Start | Exon No. | Exon Position | Exon Length | Intron No. | Intron Position | Intron Length |
|--------|----------|---------------|-------------|------------|-----------------|---------------|
| @88 bp | 78.3 | @74-180 | 106 bp | 79.1 | @181-529 | 349 bp |
| | 79.1 | @530-654 | 125 bp | 79.4 | @655-720 | 66 bp |
| | 79.4 | @721-769 | 49 bp | 79.55 | @770-875 | 105 bp |
| | 79.55 | @876-893 | 18 bp | 79.75 | @894-932 | 39 bp |
| | 79.85 | @933-966 | 33 bp | | | |

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 12- 32 1.8833

Figure 26C



Predicted TM structure

> : Too long to be significative

< : Too short to be significative

LL: Loop length

KR : Number of Lys and Arg

KR Diff : Positive charge difference

CE : Net charge energy

CE Diff : Net charge difference

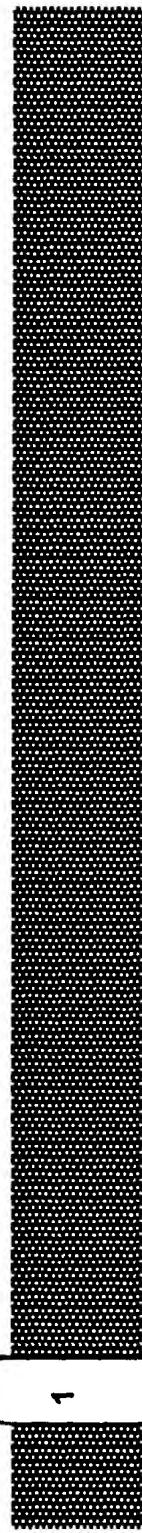
CH Diff : Charge difference over N-term segments

CE = -0.54

KR =>

LL = 75

C
N
1



N
LL = 11
KR = 1
CE = <

KR Diff = 1
CH Diff = -3
CE Diff = 0.54

CYTOPLASM
OUTSIDE
OUTSIDE

Structure no. 1

Figure 26D



A readthrough apo-4S product using the second available methionine

The Apo-4S peptide sequence

P1 Begin TM1(R)

+30 | P2

MYPIMEYSCSD RNLVLIYGIL LIYIYIGNLN VARHFSMKTP
VARSNIKLIL 80

TNNVKWLHKK GFASSWKLVK NQTLLCTPSM QLLCCLHPEM
GNDFPNGKET 130

P3

ERCYLSAPFV KSVFLSLCFP GHNVGSLFHM ADDLGRAMES
LVSVMTDEEG 180

AEKMFYNSRF PHGFYNIHTT KRIRQ**KE**FTR NKSIFL**RR**VV VLYCRF**QK**FL
230

SLLLFCK**QWQ VLHVYAI**V**QK SYKK**TT**CKIL IAKKLAISLY**
GTHFGLFKNL 280

KQLKRKNYKG KRKKRNGQVV KLRTQVCTII RNTPKPKRGR
NSMRSRV**R**CK 330

LI 332 (302aa in predicted polypeptide) (SEQ ID NO: 56)

Figure 27A



Candidate membrane-spanning segments:

| | | | |
|----------|---|---------|--------|
| Certain | 1 | 41-61 | 1.9073 |
| Putative | 2 | 101-121 | 0.8052 |
| Certain | 3 | 132-152 | 1.2552 |
| Putative | 4 | 217-237 | 1.1833 |
| Putative | 5 | 254-274 | 0.9240 |

Transmembrane segments included in structure No. 8: 1 2 3 4 5

Loop lengths: 11 39 10 64 16 58; K+R profile: 1 2 5 (9 >22)

K+R difference: -23: -> Orientation: **N-out**

Charge-difference over N-terminal Membr. segs. (± 15 residues): -4
-> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < -0.13 <

CYT-EXT difference: 0.13: -> Orientation: **N-out**

Figure 27B



> : Too long to be significative
< : Too short to be significative
LL : Loop length
KR : Number of Lys and Arg

KR Diff : Positive charge difference

CE : Net charge energy

CE Diff : Net charge difference

CH Diff : Charge difference over N-term segments

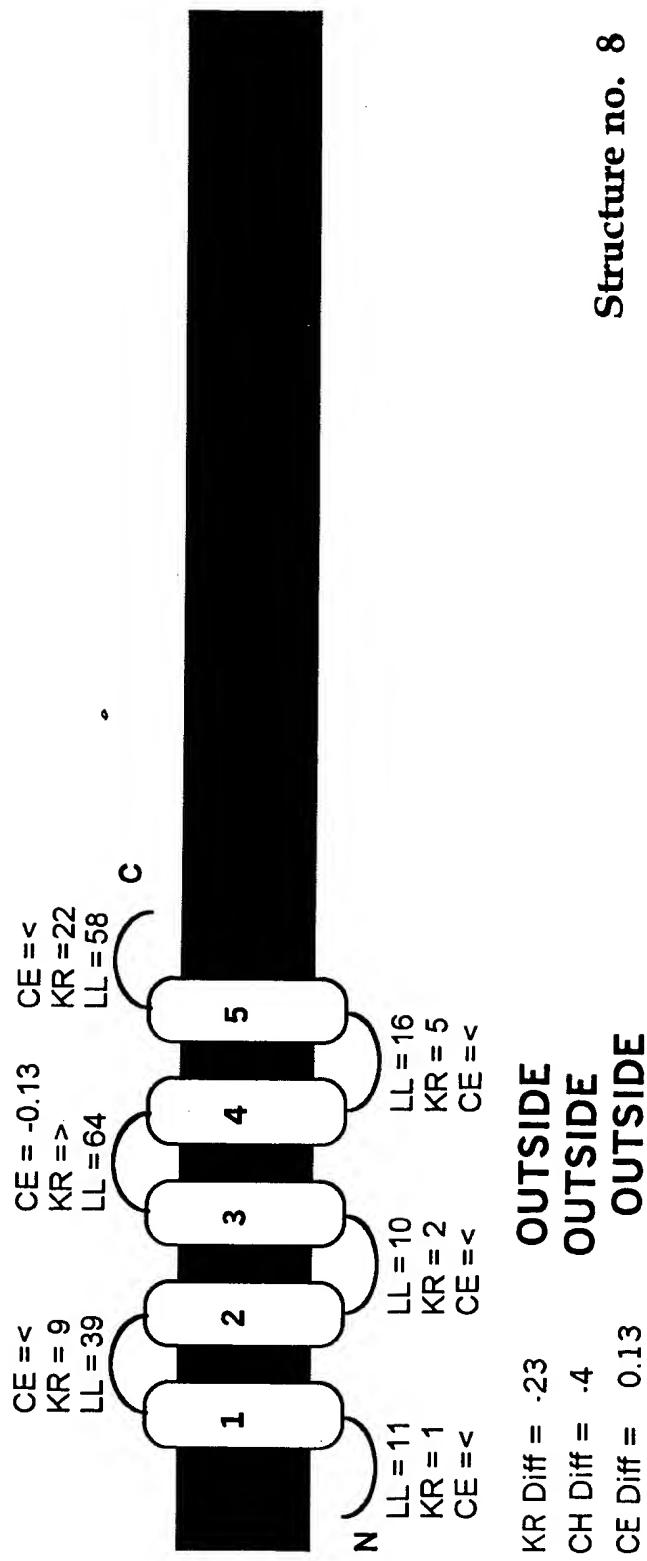




Figure 28

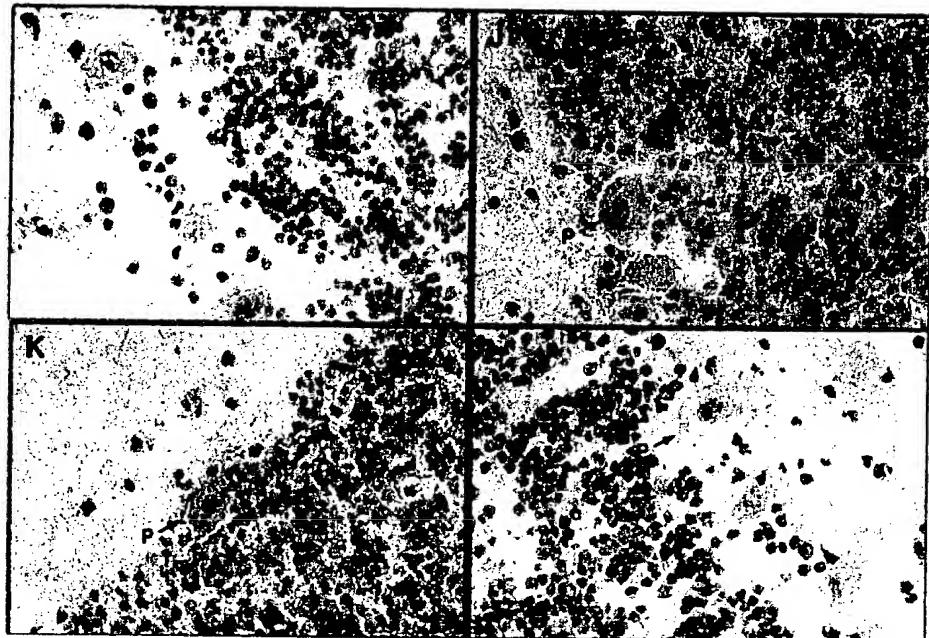


Figure 28 (cont'd)

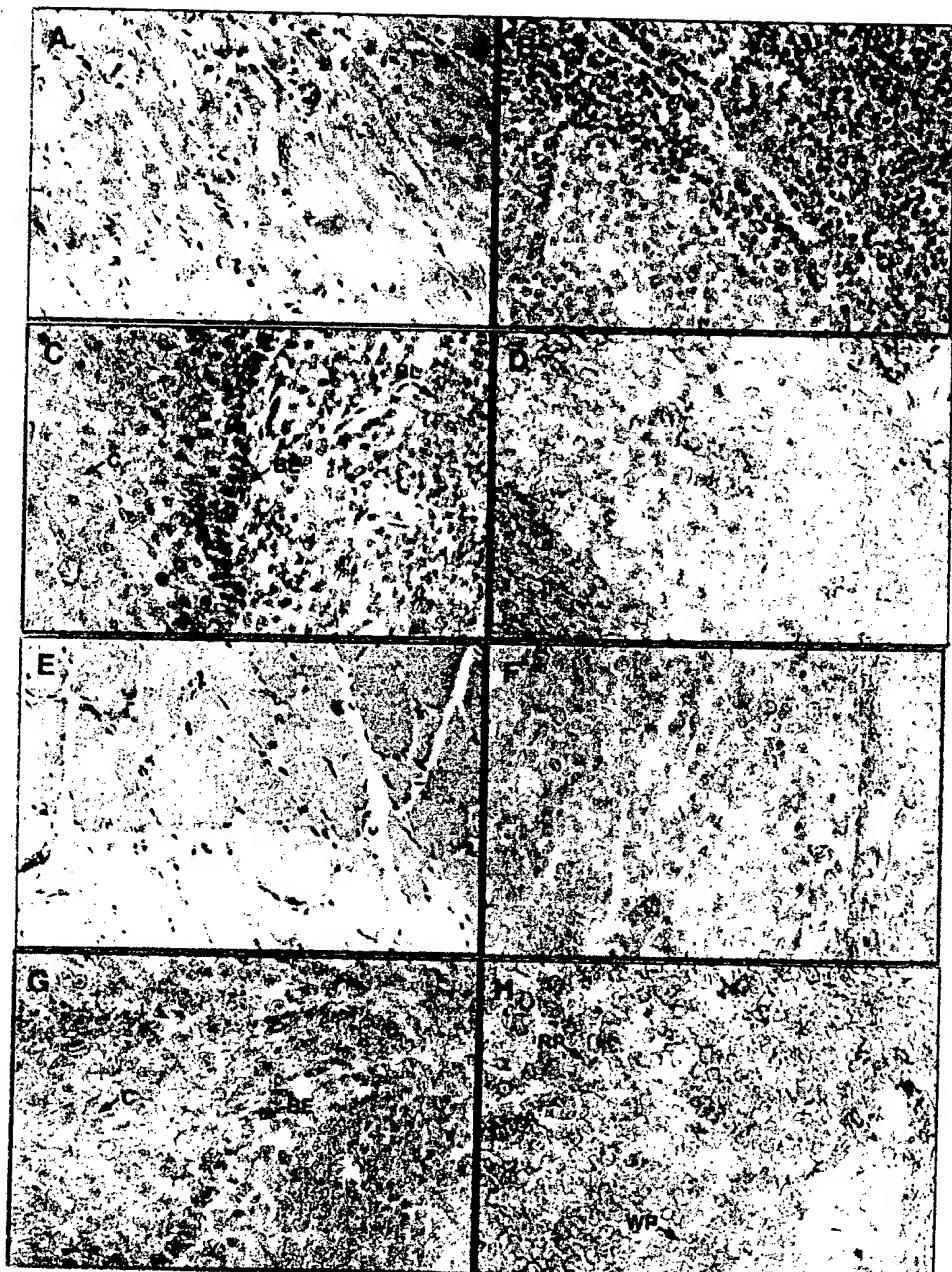


Figure 29

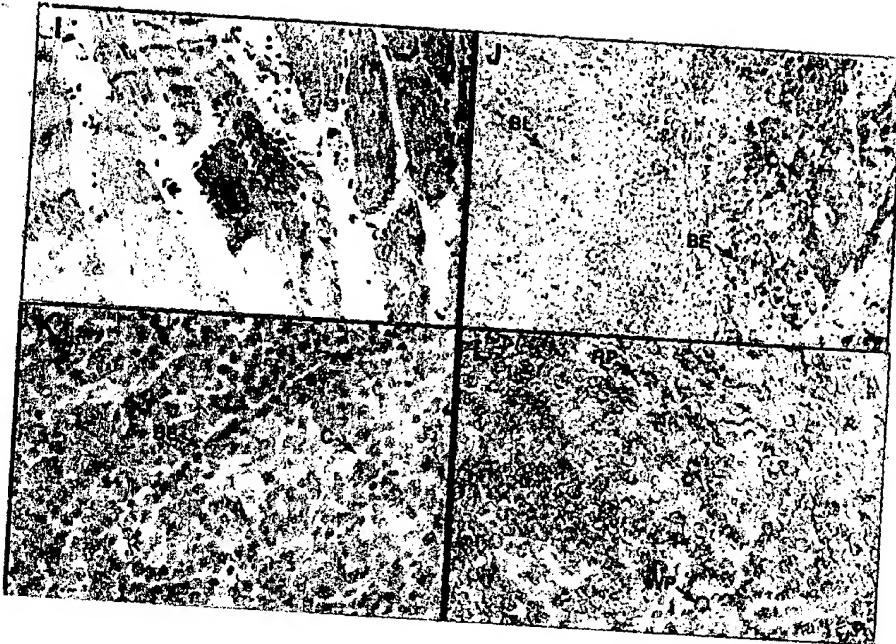


Figure 29 (cont'd)

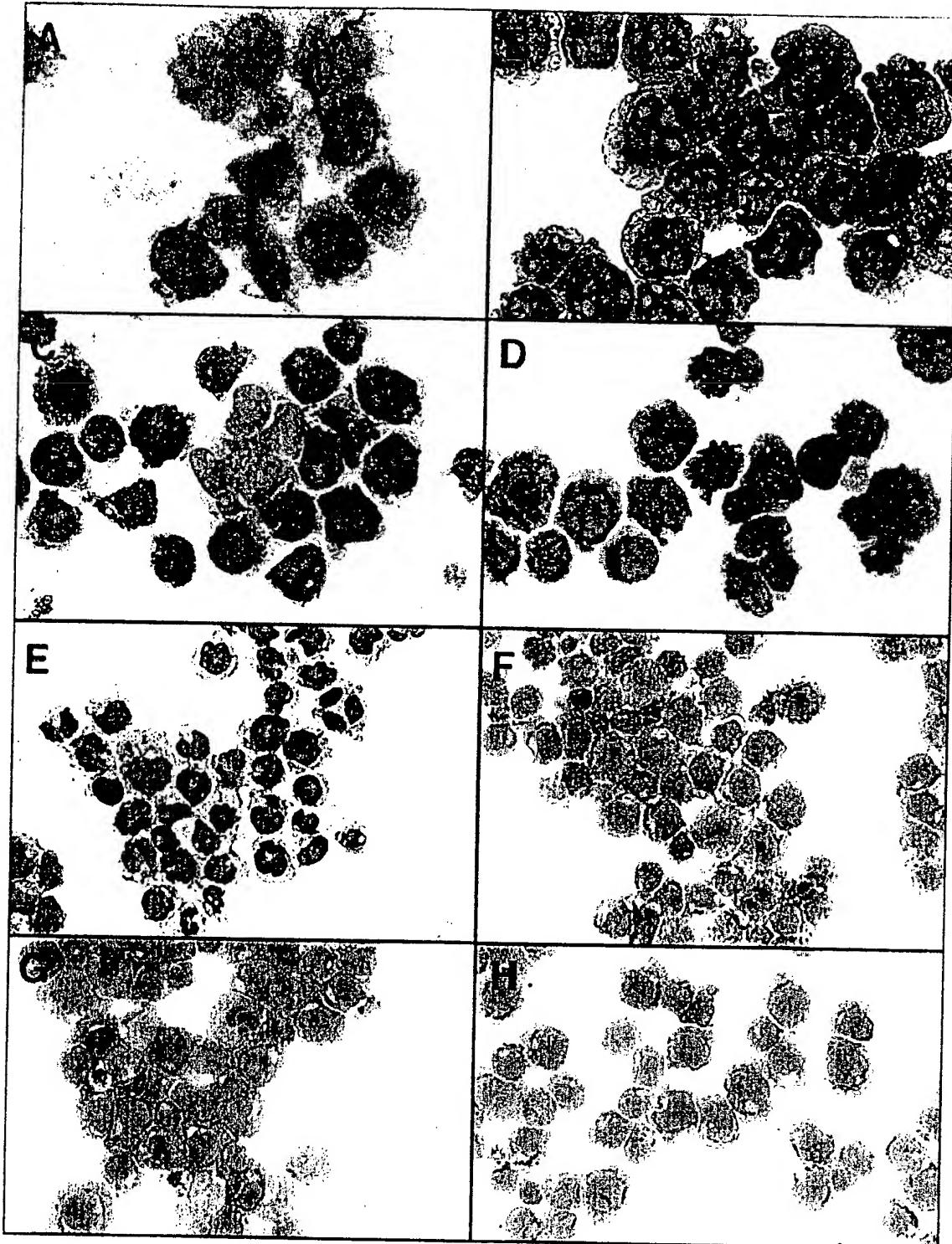


Figure 30



RECEIVED

SEP 18 2002

TECH CENTER 1600/2900

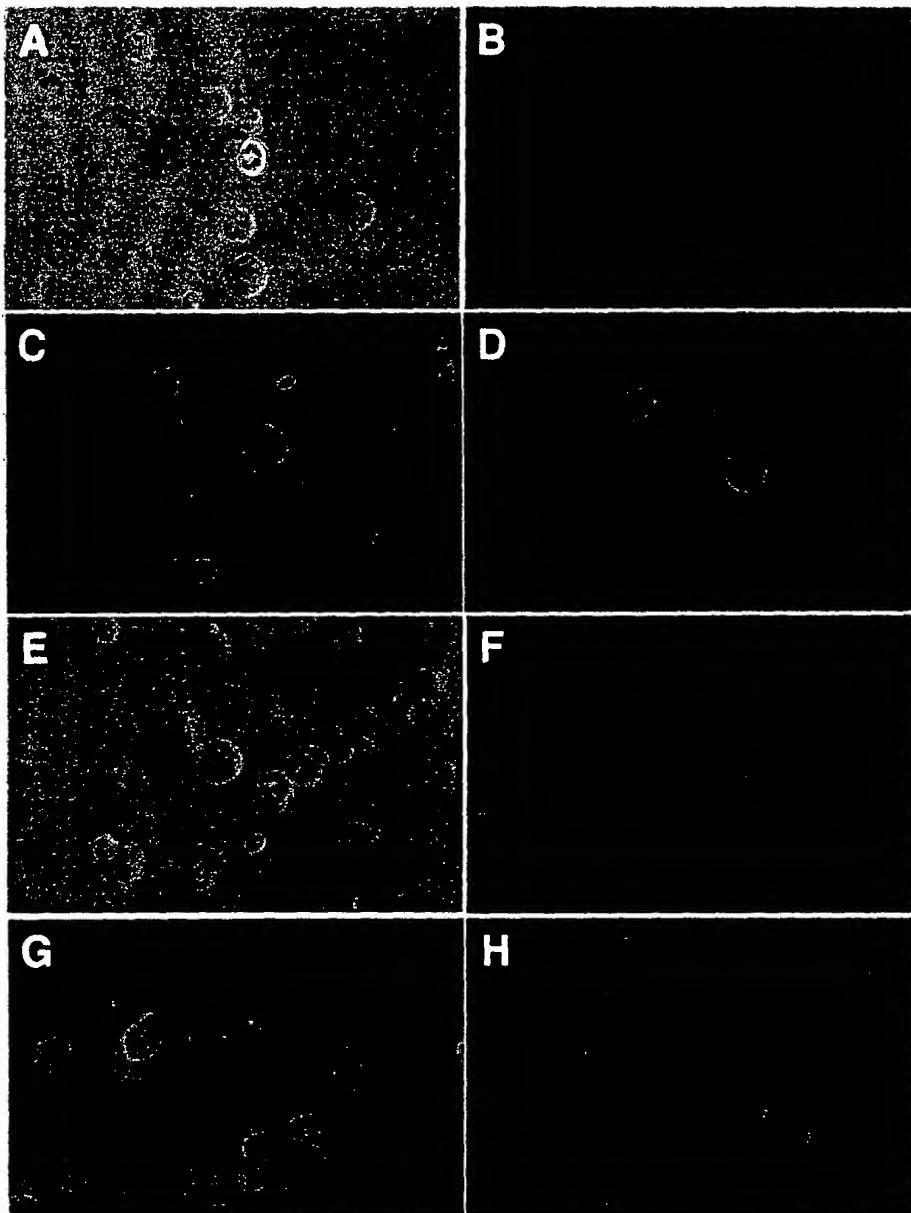


Figure 31

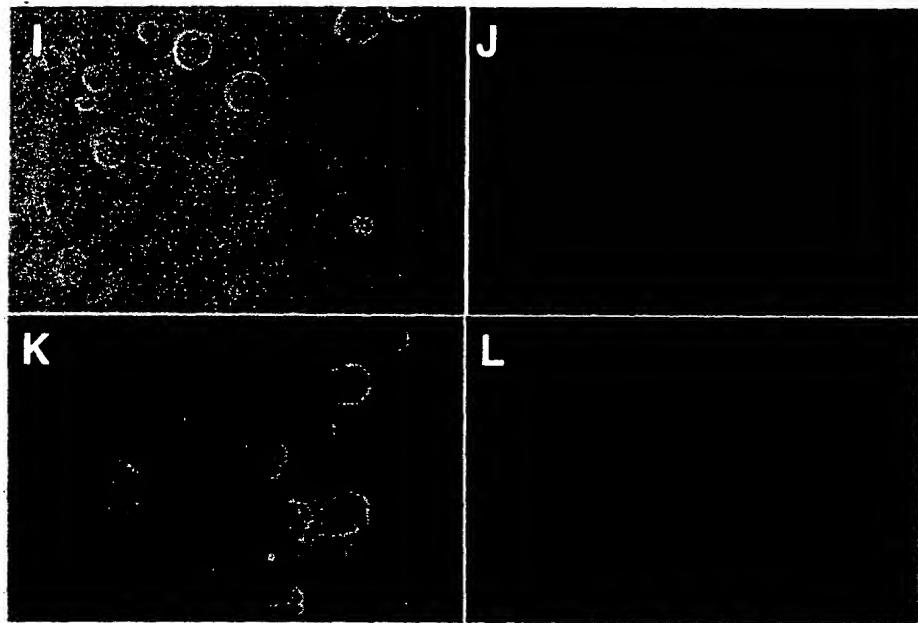


Figure 31 (cont'd)

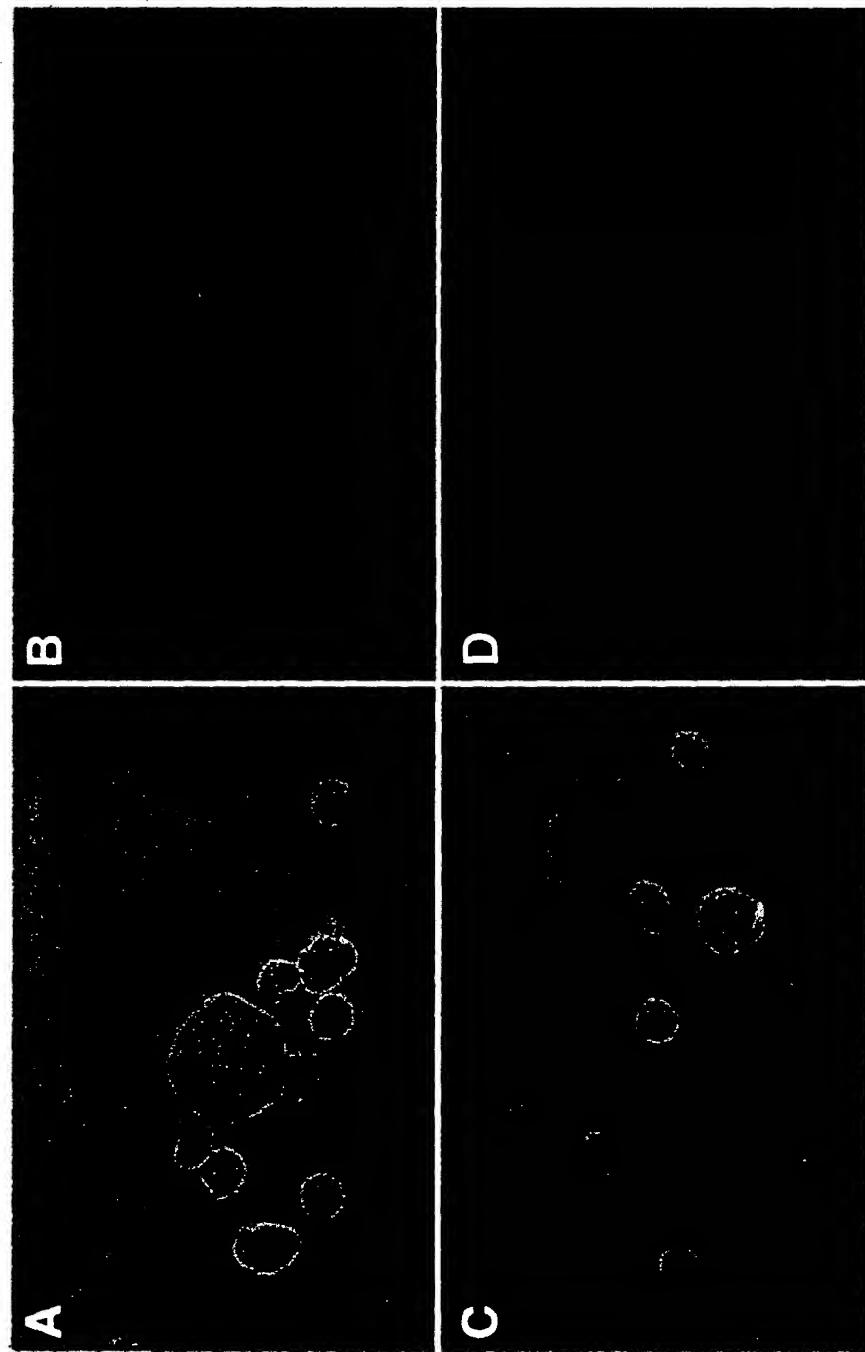


Figure 32



RECEIVED

SEP 18 2002

TECH CENTER 1600/2900

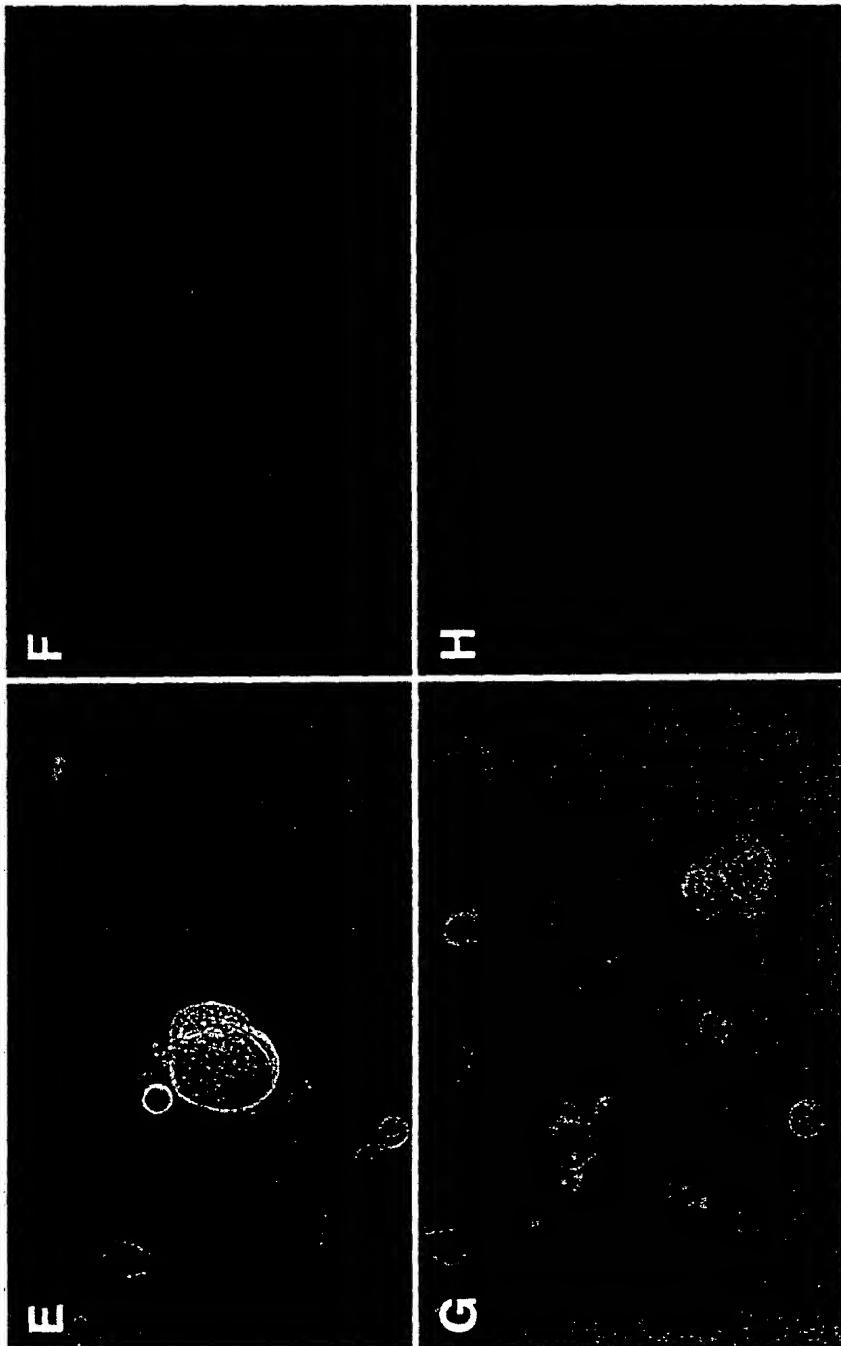


Figure 32 (cont'd)

SEP 18 2002

TECH CENTER 1600/2900

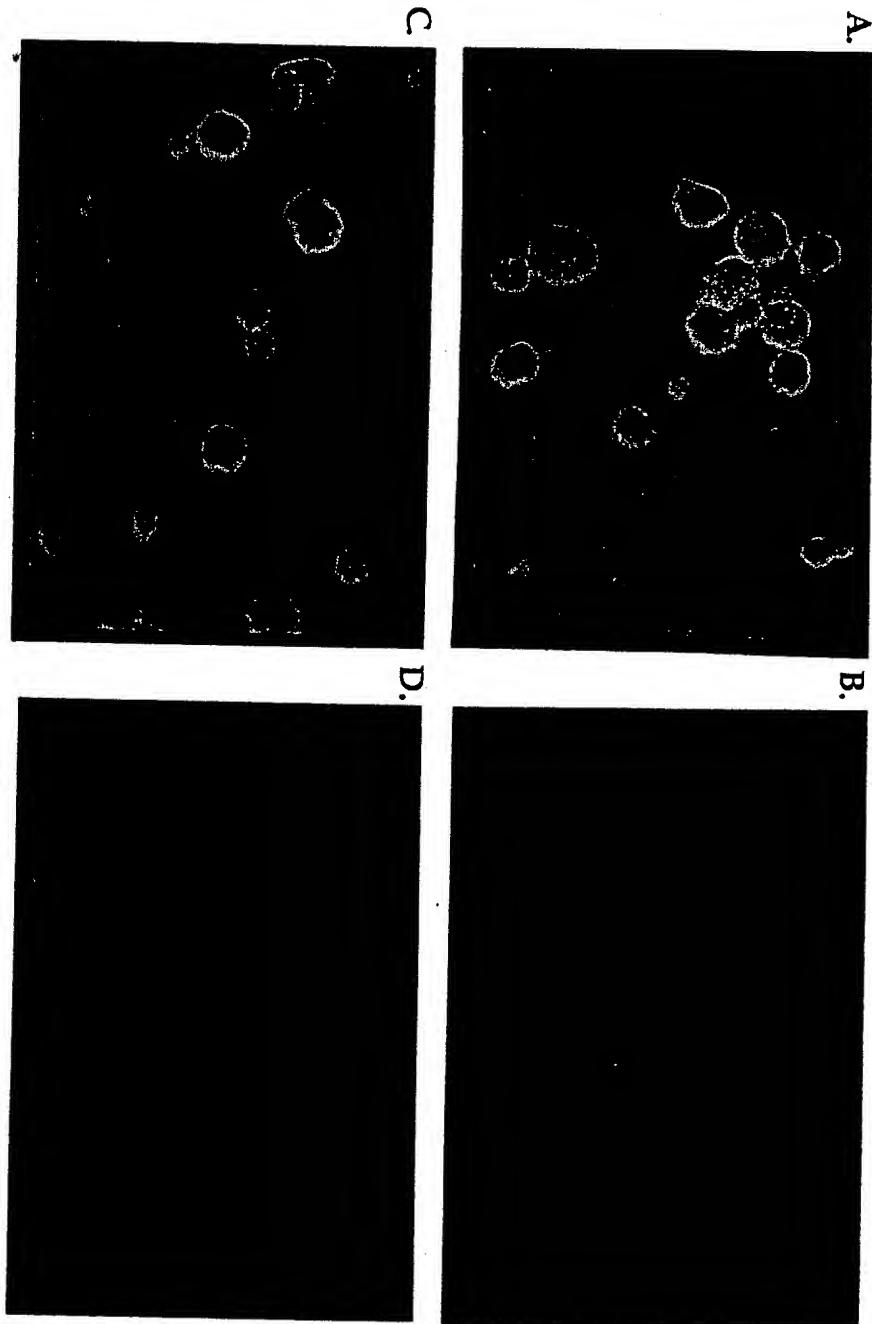


Figure 33



Additional Oligonucleotide primers used for apo-dystrophin-4 southern blotting and sequencing

FORWARD

| | | |
|--------------------------------|------------|----------------|
| GTT CGT TAA TAC AAG TAG | F2.3(@28) | (SEQ ID NO 15) |
| GCC AAG GTG GAA AAG ATG | F2.2(@73) | (SEQ ID NO 16) |
| CCA GTA GCC TGA TCC AAC | F3.2(@208) | (SEQ ID NO 17) |
| GGC TTC ATT AAT AAG | F3.1(@257) | (SEQ ID NO 18) |
| GGC AAA GAA ACA GAG TG | F4.2(@379) | (SEQ ID NO 19) |
| CAG GAC ACA ATG TAG GA | F4.1(@449) | (SEQ ID NO 20) |
| GTT ATA AAG AAA GAA TTA TAA AG | FJn(@846) | (SEQ ID NO 21) |
| GAA AAT AAC GCA ATG GAC | F5.1(@875) | (SEQ ID NO 22) |

REVERSE

| | | |
|-------------------------------|-------------|----------------|
| GAT GGG ATA CAT CTT TTC C | R6.1(@99) | (SEQ ID NO 23) |
| CAA GCT ACA TTC AGG TTC CC | F2.2R(@188) | (SEQ ID NO 24) |
| GGA CTC CAT CGC TCT GCC | R4.1(@510) | (SEQ ID NO 25) |
| GAC TTA GAA ACT ACT G | R3.4(@694) | (SEQ ID NO 26) |
| ATA GAC GTG TAA AAC CTG C | R2.1(@735) | (SEQ ID NO 27) |
| AAC TGT TAT AAA TTT TTA | RSP2(@848) | (SEQ ID NO 28) |
| CTT TTT CCT TTA TAA TTC TTT C | R2.3o(@875) | (SEQ ID NO 29) |

Figure 34



An Additional Splice Product Predicted From The Apo-4 Gene

A second potential theoretical splice product which retains exon 78.3 is shown below.

H2 p1-124 spliced product =351 bp, 117 amino acids + 10 from vector + 1 N-glycosylation site; predicted weight = 21.9 Kd

Figure 35A

Peptide Generated

MFVNNTKVEKMYPIMEYSCSDRNLVLIYGILLIYIYIGNLNMKEQNKCFTTPDSRMVFII
FIQQRGLDSKSLQEINLYFCEGFYTSMLQYKKVIRKLHKITQWTRTPQNQSEVEIA (117
amino acids) (SEQ ID NO 30)

Figure 35B

| Start | Exon No. | Exon Position | Exon Length | Intron No. | Intron Position | Intron Length |
|--------|----------|---------------|-------------|------------|-----------------|---------------|
| @26 bp | 78.1 | @16-41 | 26 bp | 78.3 | @42-74 | 33 bp |
| | 78.3 | @75-181 | 106 bp | 79.1 | @182-530 | 349 bp |
| | 79.1 | @531-655 | 125 bp | 79.4 | @656-721 | 66 bp |
| | 79.4 | @722-770 | 49 bp | 79.55 | @771-876 | 105 bp |
| | 79.55 | @877-894 | 18 bp | 79.75 | @895-933 | 39 bp |
| | 79.85 | @934-967 | 33 bp | | | |

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 22- 42 1.8833

Figure 35C

D. Predicted TM structure

> : Too long to be significative

< : Too short to be significative

LL : Loop length

KR : Number of Lys and Arg

KR Diff : Positive charge difference

CE : Net charge energy

CE Diff : Net charge difference

CH Diff : Charge difference over N-term segments

$$CE = -0.54$$

KR =>
LL = 75
CE = <

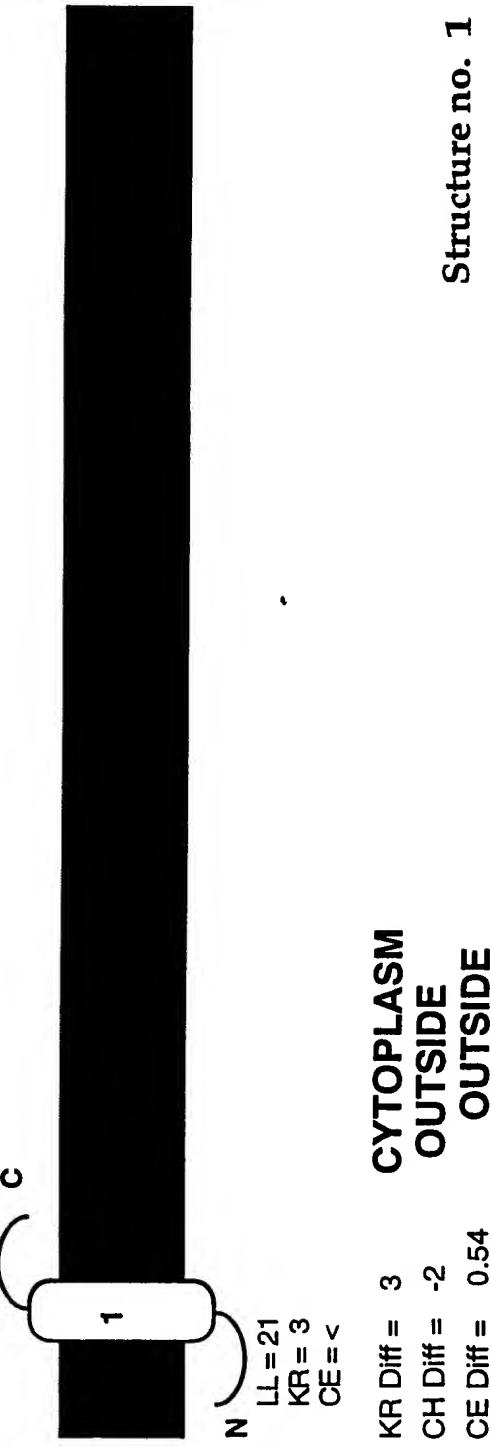


Figure 35D





Nucleic Acid Subsequence Sites Identified In Apo-4

| <u>Motif</u> | <u>Position</u> | <u>Significance</u> |
|--------------------------|--------------------|--|
| CpG | -7, (+28, +106) | DNA methylation site |
| CAAT | -132, (+127, +131) | Binding of CAAT factors |
| TATAAT (5/6) | -120, -114, (+10) | TFIID Binding site |
| TATA | -154 | Binds RNA polymerase II and TFIID |
| CCATTCA | -162, -131 | Cap Site I |
| TATCAGT | +12, (+25) | Cap Site II |
| TGGCTGCAAGCCCCAA (10/14) | -57, (+41) | Binds CTF/NF-I protein |
| (SEQ ID NO: 32) | | |
| GTGATGG | -140, -4, +11, +32 | Eucaryotic Transcription Initiation Site |

Figure 36



Top Pred predicts 4-5 transmembrane domains for a full-length apo-4F product in which all the stop codons are suppressed.

Protein sequence and position of predicted TM domains

Begin TM₁ (R)

P1 | P2

MFVNTSREKV INQSLIAKVE KMYPIMEYSCSD RNLVLIYGIL LIYIYIGNLN VARHFSMK60

TPVARSNIKL ILTNNVKWLH KKGFASSWKL VKNQTLLCTP SMQLLCLHP EMGNDFPNGK 120

P3

ETERCYLSAP FVKSVFLSLC FPGHNVGSLF HMADDLGRAM ESLVSVMTDE EGAEKMFYNS180

RFPHGFYNIH TTKRIRQKEF TRNKSIFLRR VVVLYCRFQK FLSLLLFCKQ WQVLHVYAIV 240

QKSYKKTTCK ILIAKKLAIS LYGTHFGLFK NLKQLKRKNY KGKRKKRNGQ VVKLRTQVCT 300

IIRNTPKPKR GRNSMRSRVR CKLI (324 amino acids) (SEQ ID NO 31)

Hydrophobicity Scale KD

Figure 37A

Apo-4F : Candidate membrane-spanning segments:

| | | | |
|----------|---|----------|--------|
| Certain | 1 | 33- 53 | 1.9073 |
| Putative | 2 | 93- 113 | 0.8052 |
| Certain | 3 | 124- 144 | 1.2552 |
| Putative | 4 | 209- 229 | 1.1833 |
| Putative | 5 | 246- 266 | 0.9240 |

I. Transmembrane segments included in structure 8: 1 2 3 4 5; Loop lengths: 32 39 10 64
16 58

Figure 37B



K+R difference: -19; -> Orientation: **N-out**; Charge-difference over N-terminal Membr. segs.

(±15 residues): -3; -> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < < < < -0.13 <

CYT-EXT difference: 0.13

-> Orientation: **N-out**

II. Transmembrane segments included in structure 7: **1 3 4 5**; Loop lengths: 32 70 64 16
58

K+R profile: 5 > 22 > 5; K+R difference: 22 -> Orientation: **N-in**

Charge-difference over N-terminal Membr. segs. (±15 residues): -3; -> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < -0.13 < -0.26 <

CYT-EXT difference: 0.13; -> Orientation: **N-out**

Figure 37B (cont'd)



C. TopPred predicts a cytoplasmic N-terminus for four TM domains

>: Too long to be significative
 <: Too short to be significative
 LL: Loop length
 KR: Number of Lys and Arg

KR Diff : Positive charge difference
 CE : Net charge energy
 CE Diff : Net charge difference
 CH Diff : Charge difference over N-term segments

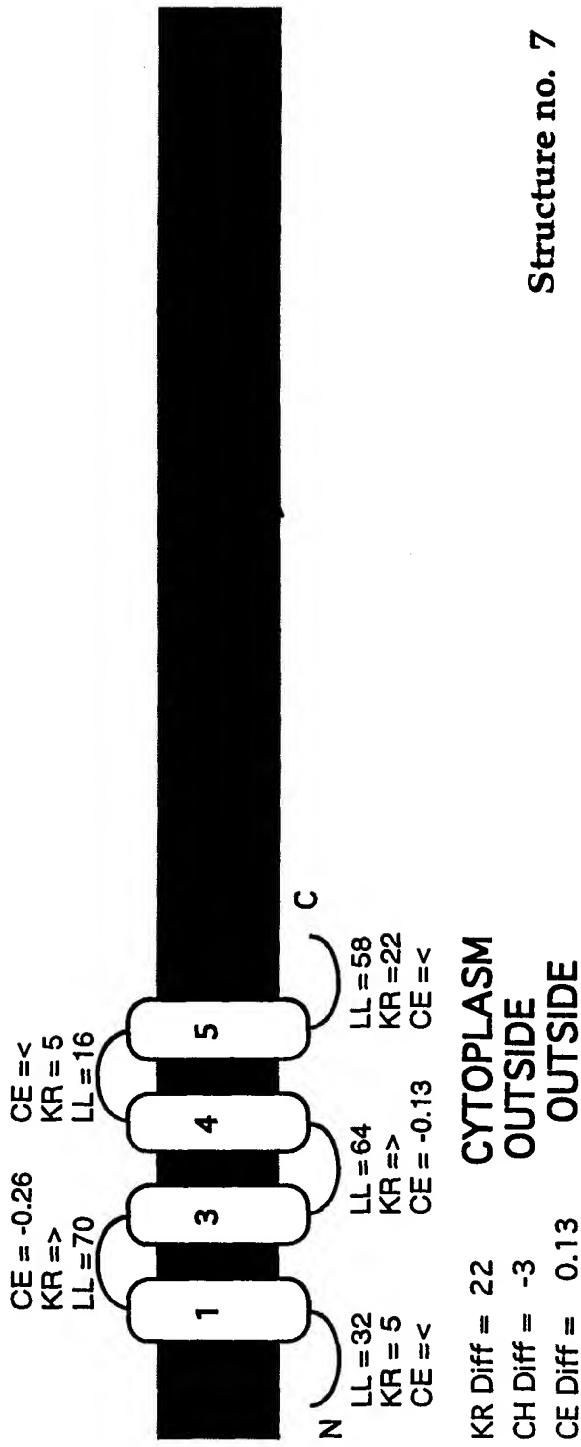


Figure 37C

Basic Features of a Transposon or Retrovirus

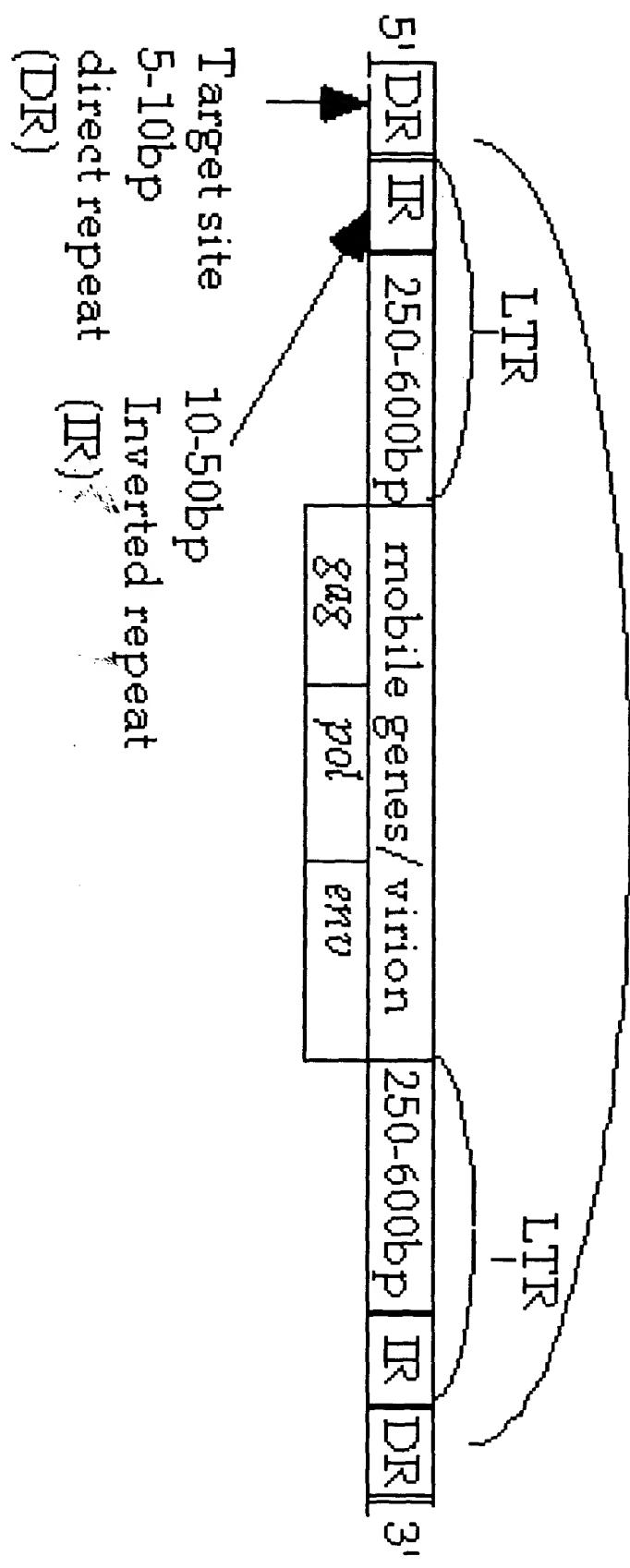


Figure 38A



Structure of the apo-4 inversion element before rearrangement

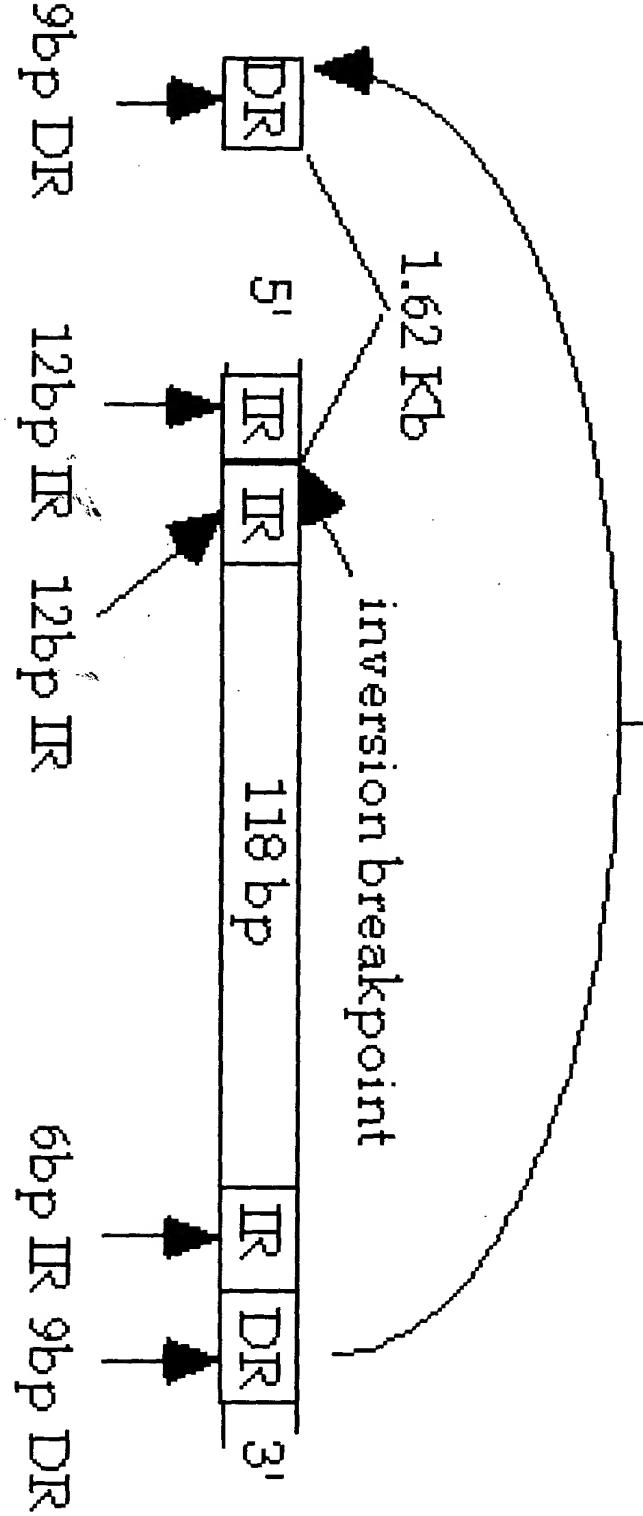
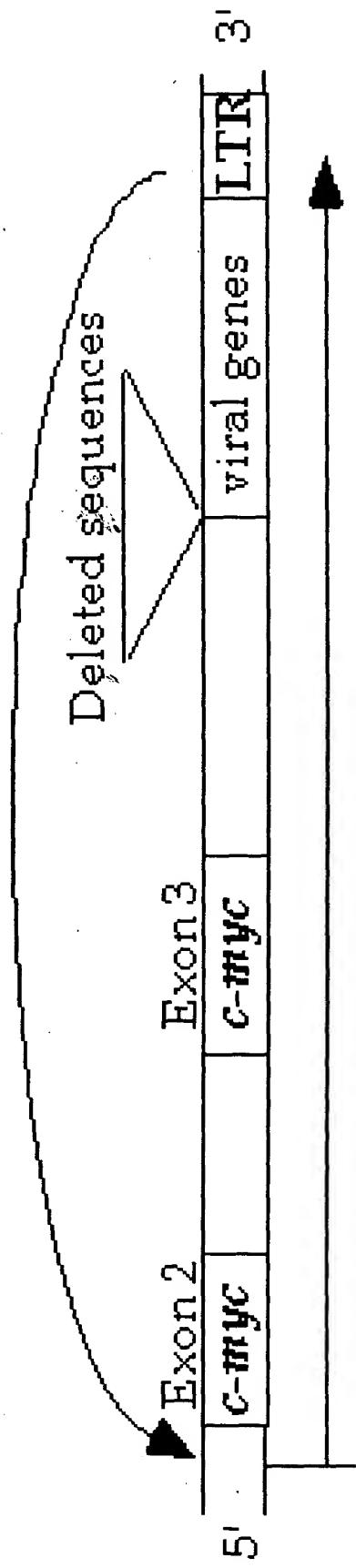


Figure 38B





RNA transcript is promoted from cell sequences but enhanced and terminated by viral sequences.

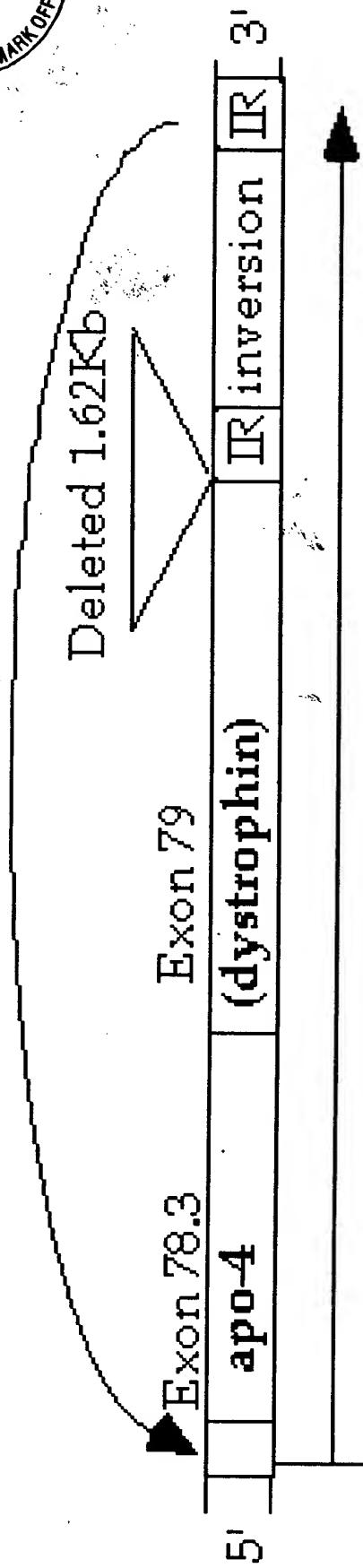
Figure 39A



RECEIVED

SEP 18 2002

TECH CENTER 1600/2900



RNA transcript is promoted from cell sequences but enhanced and terminated by inversion sequences which may also activate suppressor tRNAs or reverse transcriptase activity to prevent the recognition of stop codons. Inverted repeats (IR) are present at both ends of the inversion, as they are in retroviruses and transposable elements.

Figure 39B